Community-Level Responses to Iron Availability in Open Ocean Planktonic Ecosystems

Luigi Caputi[†1], Quentin Carradec[†2, 3, 4], Damien Eveillard[†5], Amos Kirilovsky[†6, 7], Eric Pelletier[†2, 3, 4], Juan J. Pierella Karlusich[†6], Fabio Rocha Jimenez Vieira[†6], Emilie Villar[†6, 8], Samuel Chaffron[5], Shruti Malviya[6, 9], Eleonora Scalco[1], Silvia G. Acinas[10], Adriana Alberti[2], Jean-Marc Aury[2], Anne-Sophie Benoiston[6, 11], Alexis Bertrand[2], Tristan Biard[12, 13], Lucie Bittner[6, 8, 11], Martine Boccara[6], Jennifer R. Brum[14], Christophe Brunet[1], Greta Busseni[1], Anna Carratalà[15], Hervé Claustre[13], Luis Pedro Coelho[16], Sébastien Colin[6, 8], Salvatore D'Aniello[1], Corinne Da Silva[2], Marianna Del Core[17], Hugo Doré[8], Stéphane Gasparini[13], Florian Kokoszka[1, 6, 18], Jean-Louis Jamet[19, 20], Christophe Lejeusne[1, 51], Cyrille Lepoivre[21], Magali Lescot[22], Gipsi Lima-Mendez[23, 26], Fabien Lombard[13], Julius Lukeš[27, 28], Nicolas Maillet[1, 29], Mohammed-Amin Madoui[2, 3, 4], Elodie Martinez[30], MariaGrazia Mazzocchi[1], Mario B. Néou[2, 3, 4], Javier Paz-Yepes[6], Julie Poulain[2], Simon Ramondenc[13], Jean-Baptiste Romagnan[13], Simon Roux[14], Daniela Salvagio Manta[17], Remo Sanges[1], Sabrina Speich[18], Mario Sprovieri[17], Shinichi Sunagawa[16, 31], Vincent Taillander[13], Atsuko Tanaka[6], Leila Tirichine[6], Camille Trottier[32], Julia Uitz[13], Alaguraj Veluchamy[6, 33], Jana Veselá[27], Flora Vincent[6], Sheree Yau[34], Stefanie Kandels-Lewis[16, 35], Sarah Searson[13], Céline Dimier[6, 8], Marc Picheral[13], *Tara* Oceans Coordinators, Peer Bork[35], Emmanuel Boss[36], Colomban de Vargas[8], Michael J. Follows[37], Nigel Grimsley[34], Lionel Guidi[13, 38], Pascal Hingamp[22], Eric Karsenti[6, 35], Paolo Sordino[1], Lars Stemmann[13], Matthew B. Sullivan[14], Alessandro Tagliabue[39], Adriana Zingone[1], Laurence Garczarek[8], Fabrizio d'Ortenzio[13], Pierre Testor[40], Fabrice Not[8], Maurizio Ribera d'Alcalà[*1], Patrick Wincker[*2, 3, 4], Chris Bowler[$*6], Daniele Iudicone[$*1]

Tara Oceans Coordinators: Silvia G. Acinas[10], Peer Bork[35, 41, 42, 43], Emmanuel Boss[36], Chris Bowler[$*6], Colomban de Vargas[8], Michael J. Follows[37], Gabriel Gorsky[13], Nigel Grimsley[34], Pascal Hingamp[22], Daniele Iudicone[$*1], Olivier Jaillon[2, 3], Stefanie Kandels-Lewis[16, 35], Lee Karp-Boss[36], Eric Karsenti[6, 35], Uros Krzic[44], Fabrice Not[8], Hiroyuki Ogata[45], Stéphane Pesant[46, 47], Jeroen Raes[23], Emmanuel G. Reynaud[48], Christian Sardet[13], Mike Sieracki[49, 50], Sabrina Speich[18], Lars Stemmann[13], Matthew B. Sullivan[14], Shinichi Sunagawa[16, 31], Didier Velayoudon[51], Jean Weissenbach[2, 3, 4], Patrick Wincker[*2, 3, 4]

[1]: Stazione Zoologica Anton Dohrn, Villa Comunale, 80121, Naples, Italy. [2]: CEA - Institut François Jacob, Genoscope, 2 rue Gaston Crémieux, 91057 Evry, France. [3]: CNRS UMR 8030, Evry, France. [4]: Université d'Evry Val d'Essonne, Université Paris-Saclay, Evry, France. [5]: Laboratoire des Sciences du Numérique de Nantes (LS2N) – CNRS, Université de Nantes, École Centrale de Nantes, IMT Atlantique, 2 rue de la Houssinière, 44322 Nantes, France. [6]: Institut de biologie de l'Ecole normale supérieure (IBENS), Ecole normale supérieure, CNRS, INSERM, PSL Université Paris 75005 Paris, France. [7]: INSERM, UMRS1138, Laboratory of Integrative Cancer Immunology, Centre de Recherche des Cordeliers, 15 Rue de l'Ecole de Medecine, 75006 Paris, France. [8]: CNRS, UMR 7144, Station Biology de Roof, Place Georges Teissier, 29680 Roscoff, France. [9]: Simons Centre for the Study of Living Machines, National Centre for Biological Sciences, Tata Institute of Fundamental Research, Bangalore 560065, India. [10]: Department of Marine Biology and Oceanography, Institute of Marine Sciences (ICM)-CSIC, Pg. Marítim de la Barceloneta 37-49, Barcelona E0800, Spain. [11]: Sorbonne Universités, UPMC Univ Paris 06, Univ Antilles Guyane, Univ Nice Sophia Antipolis, CNRS, Evolution Paris Seine - Institut de Biologie Paris Seine (EPS - IBPS), 75005 Paris, France. [12]: Sorbonne Universités, UPMC Université Paris 06, CNRS, Laboratoire Adaptation et Diversitéen Milieu Marin UMR7144, Station Biologique de Roscoff, 29688

Roscoff, France. [13]: Sorbonne Universités, UPMC Université Paris 06, CNRS, Laboratoire d'Océanographie de Villefranche (LOV) UMR7093, Observatoire Océanologique, 06230 Villefranche-sur-Mer, France. [14:] Departments of Microbiology and Civil, Environmental and Geodetic Engineering, The Ohio State University, Columbus OH 43210, USA. [15]: Laboratory of Environmental Chemistry, School of Architecture, Civil and Environmental Engineering (ENAC), École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland. [16]: Structural and Computational Biology Unit, European Molecular Biology Laboratory, 69117 Heidelberg, Germany. [17]: Institute for Coastal Marine Environment (IAMC) - CNR of Capo Granitola, Via del Mare 3, Torretta Granitola (TP), 91021, Italy. [18]: LMD Laboratoire de météorologiedynamique. Ecole normale supérieure de Paris, PSL Research University, 24 rue Lhomond, 75231 Paris Cedex 05, France. [19]: Université de Toulon, Laboratoire PROTEE-EBMA E. A. 3819, BP 20132, 83957 La GardeCedex, France. [20]: Université de Toulon, Aix Marseille Universités, CNRS/INSU, IRD, MIO UM 110 Mediterranean Institut of Oceanography, La Garde, France. [21]: Information Génomique et Structurale, UMR7256, CNRS, Aix-Marseille Université, Institut de Microbiologie de la Méditerranée (FR3479), ParcScientifique de Luminy, Marseille, France. [22]: Aix Marseille Univ, Université de Toulon, CNRS, IRD, MIO, Marseille, France. [23]: Department of Microbiology and Immunology, Rega Institute, KU Leuven, Herestraat 49, 3000 Leuven, Belgium. [24]: Center for the Biology of Disease, VIB, Herestraat 49, 3000 Leuven, Belgium. [25]: Cellular and Molecular Microbiology , Faculté des Sciences, Université Libre de Bruxelles (ULB), Belgium. [26]: Interuniversity Institute for Bioinformatics in Brussels (IB2), ULB-VUB, Boulevard du Triomphe CP 263, 1050 Brussels, Belgium. [27]: Institute of Parasitology, Biology Centre CAS, Branišovská 1160/31, 370 05 ČeskéBudějovice, Czech Republic. [28]: Faculty of Science, University of South Bohemia, 370 05 Ceské Budejovice, Czech Republic. [29]: Institut Pasteur - Bioinformatics and Biostatistics Hub - C3BI, USR 3756 IP CNRS - Paris, France. [30]: IRD - BP 529 - 98713 Papeete, Tahiti. [31]: Institute of Microbiology, Department of Biology, ETH Zurich, Vladimir-Prelog-Weg 4, 8093 Zurich, Switzerland. [32]: IFREMER, Physiology and Biotechnology of Algae Laboratory, rue de l'Iled'Yeu, 44311, Nantes, France. [33]: Biological and Environmental Sciences and Engineering Division, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia. [34]: Sorbonne Universités, UPMC Univ Paris 06, CNRS, Biologie Intégrative des Organismes Marins (BIOM, UMR 7232), Observatoire Océanologique, Banyuls sur Mer, France. [35]: Directors' Research European Molecular Biology Laboratory Meyerhofstr. 1, 69117 Heidelberg, Germany. [36]: School of Marine Sciences, University of Maine, Orono, Maine, USA. Dept of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [37]: Dept of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [38]: Department of Oceanography, University of Hawaii, Honolulu, Hawaii 96822, USA. [39]: Department of Earth Ocean and Ecological Sciences, School of Environmental Sciences, University of Liverpool, Liverpool, UK. [40]: Sorbonne Universités (UPMC, Univ Paris 06)-CNRS-IRD-MNHN, Laboratoire LOCEAN, Paris, France. [41]: Structural and Computational Biology, European Molecular Biology Laboratory, Meyerhofstr. 1, 69117 Heidelberg, Germany. [42]: Max Delbrück Centre for Molecular Medicine, 13125 Berlin, Germany. [43]: Department of Bioinformatics, Biocenter, University of Würzburg, 97074 Würzburg, Germany. [44]: Cell Biology and Biophysics, European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany. [45]: Institute for Chemical Research, Kyoto University, Gokasho, Uji, Kyoto, 611-0011, Japan. [46]: MARUM, Center for Marine Environmental Sciences, University of Bremen, Bremen, Germany. [47]: PANGAEA, Data Publisher for Earth and Environmental Science, University of Bremen, Bremen, Germany. [48]: Earth Institute, University College Dublin, Dublin, Ireland. [49] National Science Foundation, Arlington, VA 22230, USA. [50]: Bigelow Laboratory for Ocean Sciences East Boothbay, ME, USA. [51]: DVIP Consulting, Sèvres, France. . [51:] Institut Méditerranéen de Biodiversité et d'Ecologie Marine et Continentale (IMBE), UMR 7263 CNRS, IRD, Aix Marseille Université, Avignon Université, Station Marine d'Endoume, Marseille, France.

* Correspondence to: pwincker@genoscope.cns.fr, cbowler@biologie.ens.fr, iudicone@szn.it, maurizio@szn.it

† These authors contributed equally to this work

$ Equal coordinating contribution

Key Points:

- Coherent assemblages of taxa co-varying with iron at global level are identified in plankton communities
- Functional responses to iron availability involve both changes in copy numbers of iron-responsive genes and their transcriptional regulation
- Plankton responses to local variations in iron concentrations recapitulate global patterns

Abstract

Predicting responses of plankton to variations in essential nutrients is hampered by limited *in situ* measurements, a poor understanding of community composition, and the lack of reference gene catalogs for key taxa. Iron is a key driver of plankton dynamics and, therefore, of global biogeochemical cycles and climate. To assess the impact of iron availability on plankton communities we explored the comprehensive bio-oceanographic and -omics datasets from *Tara* Oceans in the context of the iron products from two state-of-the-art global scale biogeochemical models. We obtained novel information about adaptation and acclimation towards iron in a range of phytoplankton, including picocyanobacteria and diatoms, and identified whole sub-communities co-varying with iron. Many of the observed global patterns were recapitulated in the Marquesas archipelago, where frequent plankton blooms are believed to be caused by natural iron fertilization, although they are not captured in large scale biogeochemical models. This work provides a proof-of-concept that integrative analyses, spanning from genes to ecosystems and viruses to zooplankton, can disentangle the complexity of plankton communities and can lead to more accurate formulations of resource bioavailability in biogeochemical models, thus improving our understanding of plankton resilience in a changing environment.

Plain Language Summary

Marine phytoplankton require iron for their growth and proliferation. According to John Martin's iron hypothesis, fertilizing the ocean with iron could dramatically increase photosynthetic activity, thus representing a biological means to counteract global warming. However, while there is a constantly growing knowledge of how iron is distributed in the ocean and about its role in cellular processes in marine photosynthetic groups such as diatoms and cyanobacteria, less is known about how iron availability shapes plankton communities and how they respond to it. In the present work, we exploited recently published *Tara* Oceans datasets to address these questions. We firstly defined specific subcommunities of co-occurring organisms that are directly related to iron availability in the oceans. We then identified specific patterns of adaptation and acclimation to iron in different groups of phytoplankton. Finally, we validated our global results at local scale, specifically in the Marquesas archipelago, where recurrent iron-driven phytoplankton blooms are believed to be a result of iron fertilization. By integrating

global data with a localized response we provide a framework for understanding the resilience of plankton ecosystems in a changing environment.

1. Introduction

Marine plankton play critical roles in pelagic oceanic ecosystems. Their photosynthetic component (phytoplankton, consisting of eukaryotic phytoplankton and cyanobacteria) accounts for approximately half of Earth's net primary production, fueling marine food webs and sequestration of organic carbon to the ocean interior. Phytoplankton stocks depend on the availability of primary resources such as nutrients which are characteristically limiting in the oligotrophic ocean. For example, high nutrient low chlorophyll (HNLC) regions are often lacking the key micronutrient iron, and increased bioavailability of iron will typically trigger a phytoplankton bloom *(Boyd et al., 2007).* Notwithstanding, the community response and its impact on food web structure and biogeochemical cycles are seldom predictable. The composition of blooms when limiting nutrients are supplied as sudden pulses with respect to the pre-existing community has been only poorly explored, and is even more elusive when comparing to situations when nutrients are in quasi-steady-state. Characterizing these responses is crucial to anticipate future changes in the ocean yet is challenged by community complexity and processes that span from genes to ecosystems. Dissecting these processes would also enhance the robustness of existing biogeochemical models and improve their predictive power (Stec et al., 2017).

In this report we explore the responsiveness of plankton communities to iron and assess the representation of iron bioavailability in biogeochemical models. Using global comprehensive metagenomics and metatranscriptomics data from *Tara* Oceans (Guidi et al., 2016; Bork et al., 2015; Alberti et al., 2017; Carradec et al., 2018), we examine abundance and expression profiles of iron-responsive genes in diatoms and other phytoplankton, together with clade composition in picocyanobacteria and the occurrence of iron-binding sites in bacteriophage structural proteins. These profiles are compared in the global ocean with the iron products from two state-of-the-art biogeochemical models. We further identify coherent sub-communities of taxa co-varying with iron in the open ocean that we denote iron-associated assemblages (IAAs). Overall, our findings are congruent with the outputs from the models and reveal a range of adaptive and acclimatory strategies to cope with iron availability within plankton communities. As a further proof-of-

concept, we track community composition and gene expression changes within localized blooms downstream of the Marquesas archipelago in the equatorial Pacific Ocean, where previous observations have suggested them to be triggered by iron (Martinez and Maamaatuaiahutapu, 2004), even though the biogeochemical models currently lack the resolution to detect the phenomenon. Our results indicate that iron does indeed drive the increased productivity in this area, suggesting that a pulse of the resource can elicit a response mimicking global steady-state patterns.

2. Materials and Methods

2.1. Iron concentration estimates

Due to the sparse availability of direct observations of iron in the surface ocean, iron concentrations were derived from two independent global ocean simulations. The first is the ECCO2-DARWIN ocean model configured with 18 km horizontal resolution and a biogeochemical simulation which resolves the cycles of nitrogen, phosphorus, iron and silicon (Menemenlis et al., 2008). The simulation resolves 78 virtual phytoplankton phenotypes. The biogeochemical parameterizations, including iron, are detailed in (Follows et al., 2007). In brief, iron is consumed by primary producers and exported from the surface in dissolved and particulate organic form. Remineralization fuels a pool of total dissolved iron which is partitioned between free iron and complexed iron, with a fixed concentration and conditional stability of organic ligand. Scavenging is assumed to affect only free iron but all dissolved forms are bioavailable. Atmospheric deposition of iron was imposed using monthly fluxes from the model of (Mahowald et al., 2005).

PISCES (Aumont et al., 2015) is a more complex global ocean biogeochemical model than ECCO2-DARWIN, representing two phytoplankton groups, two zooplankton grazers, two particulate size classes, dissolved inorganic carbon, dissolved organic carbon, oxygen and alkalinity, as well as nitrate, phosphate, silicic acid, ammonium and iron as limiting nutrients. In brief, PISCES accounts for iron inputs from dust, sediments, rivers, sea ice and continental margins, and flexible Michaelis-Menten-based phytoplankton uptake kinetics result in dynamically varying iron stoichiometry and drives variable recycling by zooplankton and bacterial activity. Iron loss accounts for scavenging onto sinking particles as a function of a prognostic iron ligand model, dissolved iron levels and the concentration of particles. Iron loss

from colloidal coagulation is also included and accounts for both turbulent and Brownian interactions of colloids. The PISCES iron cycle we use is denoted as 'PISCES2' (Tagliabue et al., 2016) performed at the upper end of a recent inter-comparison of 13 global ocean models that included iron.

2.2. *In situ* data

To generate a limited dataset of observed dissolved iron concentrations for this analysis, we used a dissolved iron database updated from (Tagliabue et al., 2012). For this we searched for the nearest available observation at the same depth as the *Tara* Oceans sampling and collected data that was within a horizontal radius of 2 degrees from the sampling coordinates.

2.3. Marquesas archipelago sampling

Four stations within the Marquesas archipelago were sampled during the *Tara* Oceans expedition in August 2011 (Karsenti et al., 2011) using protocols described in (Pesant et al., 2015): they were denoted TARA_122, TARA_123, TARA_124, and TARA_125. The sample details and physicochemical parameters recorded during the cruise are available at PANGAEA (http://www.pangaea.de), and nucleotide data are accessible at the ENA archive (http://www.ebi.ac.uk/ena/) (see further details below).

The study was initiated by releasing a glider which characterized the water column until the end of the experiment. Firstly, the mapping of the water column structure via real-time analysis of glider data was conducted. After this initial step, the continuous inspection of near real-time satellite color chlorophyll images and altimetric data revealed a highly turbulent environment, with a mixed layer up to 100 m deep and strong lateral shearing, especially downstream of the islands, which generated an area of recirculation in the wake of the main island (Nuku Hiva). A series of four sampling stations was then planned and executed by performing the full set of measurements and sampling using the *Tara* Oceans holistic protocol (Pesant et al., 2015). Station TARA_122 sampled the HNLC pre-bloom waters upstream of the islands and thus served as a reference station for the others. This station was located 27 km upstream of the island of Nuku Hiva.

2.4. Oceanographic observations

The Biogeochemical Argo (BioArgo) float deployed in the framework of the Marquesas study (WMO 6900985) was a PROVBIO-1 free-drifter profiler (Xing et al., 2012). It was based on the "PROVOR-CTS3" model, equipped with a standard CTD sensor (to retrieve temperature and salinity parameters) together with bio-optical sensors for the estimation of chlorophyll-*a* concentrations, Coloured Dissolved Organic Matter (CDOM) and backscatter at 700 nm. It was also equipped with a radiometric sensor to estimate spectral downward irradiance at three wavelengths (412, 490 and 555 nm) and with a beam transmissometer. The data processing is discussed in (Xing et al., 2012). The profiling float was programmed to adopt a modified standard Argo strategy (Freeland et al., 2005). After deployment, it navigated at 700 m depth, to a daily maximum of 1000 m and then surfaced a first time, generally early in the morning. It then submerged again to a depth up to 400 m, to again reach the surface approximately at noon. A third profile to 400 m, followed by a subsequent resurfacing, was performed at the end of the day. During all the ascending phases, a complete profile of all the available parameters was collected. At surface, the obtained data were transmitted to land through a satellite connection and the profiler descended again to 1000 m to start another cycle. The BioArgo was deployed on-site at Station TARA_123 on 2nd August 2011. It performed 55 profiles in the Marquesas region, before moving westward in early October (then outside the study area), and then southward. It definitively ceased to function in December 2012, approximately 400 km south of the Marquesas islands and after collecting more than 150 profiles.

An autonomous glider was also deployed in the study area. A complete description of glider technology and functioning is available in (Testor et al., 2010). This glider was able to reach 1000 m depths. It was equipped with temperature and salinity sensors, an optode for oxygen concentration measurements, two Wetlab ecopucks with two fluorometers for chlorophyll and CDOM concentrations, and three backscatterometers to estimate backscatter coefficients at three wavelengths (532, 700 and 880 nm). The glider was deployed on 16th July 2011 (approximately one month before *TARA* arrived in the Marquesas archipelago), close to the position of Station TARA_122. It was recovered on 5th August 2011 by *TARA* because a malfunction in the tail

rudder had been detected. It performed approximately 250 profiles, with 35 dives at 1000 m depths and 90 dives at 500 m depths.

Analysis of trace metals was performed exclusively at the Marquesas Islands sampling stations (Stations TARA_122-TARA_125) following the methods reported in (Scelfo, 1997). Dissolved iron was not measured due to lack of technical resources.

2.5. Network analysis and correlations with iron

A co-occurrence network analysis similar to that reported in (Guidi et al., 2016) was performed to delineate feature subnetworks of prokaryotic and eukaryotic lineages, as well as viral populations, based on their relative abundance. All procedures were applied on 103 sampling sites (Guidi et al., 2016) after excluding outliers (Stations TARA_82, TARA_84 and TARA_85) on Hellinger-transformed log-scaled abundances. Computations were carried out using the R package WGCNA (Langfelder and Horvath, 2007). After building a co-occurrence weighted graph, a hierarchical clustering was performed. This resulted in the definition of several subnetworks or modules, each represented by its first principal component, called *module eigen value (ME)*. Associations between the calculated subnetworks and a given trait were measured by the pairwise Pearson correlation coefficients, as well as with corresponding *p* values corrected for multiple testing using the Benjamini & Hochberg FDR procedure, between the considered environmental trait and their respective principal components. The results are reported in the first 10 columns of the heatmap in Figure S1a. The subnetworks that showed the highest correlation scores are of interest to emphasize a putative community associated with a given environmental trait. In addition to the multiple environmental parameters previously reported (Guidi et al., 2016) we simulated iron bioavailability in *Tara* Oceans stations based on the two different models of iron concentration in the global oceans: the ECCO2-DARWIN model (Menemenlis et al., 2008) and the PISCES2 model (Aumont et al., 2015). Both models performed well in the recent global iron model intercomparison project (Tagliabue et al., 2016) and so we conducted an assessment of model outputs at *Tara* sampling locations using compilations of iron observations (Tagliabue et al., 2016) augmented by those from the GEOTRACES program (Mawji et al., 2014). ECCO2-DARWIN-derived estimates (57 stations at surface) and PISCES2 model (83 stations at surface, 44 of which also at maximum chlorophyll depth) can be found in Table S1a.

For further details on the models and for a comparison of the two, see Supplementary Material. We then identified eukaryotic, prokaryotic and viral subnetworks that correlated most strongly with iron bioavailability, denoted Iron-Associated Assemblages (IAAs). Four IAAs consisting of eukaryotic metabarcodes (de Vargas et al., 2015) were significantly associated with iron. Similarly, four viral IAAs could be identified by analysis of viral communities. Based on taxonomy, no prokaryotic IAAs with significance could be identified, however when considering prokaryotic genes (as described in (Guidi et al., 2016)), five subnetworks of prokaryotic genes could be identified.

In addition to the network analyses, we examined whether the identified subnetworks can be used as predictors of iron bioavailability. Following the protocol described in (Guidi et al., 2016), we used *Partial least square (PLS) regression*, which is a dimensionality-reduction method that aims to determine predictor combinations with maximum covariance with the *response variable*. The predictors were ranked according to their *value importance in projection* (*VIP*) using the R package *pls* (Mevik and Wehrens, 2007). For each eukaryotic IAA, their relative contribution to each sample was estimated by computing the first eigen value.

2.6. Taxonomy determinations

Taxonomic studies were performed using various methods (photosynthetic pigments, flow cytometry, optical microscopy for phyto- and zooplankton as detailed in (Villar et al., 2015); phytoplankton counts using unfiltered bottles or nets as described in (Malviya et al., 2016) and (Villar et al., 2015); mesozooplankton samples collected by vertical tows with a WP2 net (200 µm mesh aperture) from 100 m depth to the surface during the day, followed by fixation in buffered formaldehyde (2–4% final concentration) and later analyzed in the laboratory. Data from an Underwater Vision Profiler (UVP) were used to determine particle concentrations and size distributions >100 µm, (Campbell et al., 1994). To have an estimate of biomass variations in the different compartments at the Marquesas Islands, we applied empirical formulas to transform Chlorophyll-*a* (phytoplankton) or body measurements (zooplankton) to biomass. To this purpose, we used the ratio of phytoplankton biomass to Chlorophyll-*a* (Phyto C: Chl a) in the euphotic zone as previously estimated in an area with similar biogeochemical features (Campbell et al., 1994 – see Table S1b). Of note, our aim was not to determine absolute biomass but to

estimate variations in biomass between the Marquesas Islands stations. To estimate the total phytoplankton biomass, Chl a concentration from HPLC data was thus used. The relative contribution of micro-, nano- and pico- plankton to the $[Chla]_{tot}$ was estimated according to (Uitz et al., 2006). The biomass of large zooplankton was estimated using previously published conversion factors from body length to carbon content (C:L) in selected zooplankton lineages. Individual body measures were estimated from literature considering similar community composition, with the exception of the Copepoda prosome length (PL), which was herein measured. Zooscan (Bongo net, >300 µm) derived abundance data (ind x m^{-3}) were used to evaluate the total biomass along the water column.

2.7. Genomic analyses

Eukaryotes larger than 5 µm were collected directly from the ocean using nets with different mesh sizes while smaller organisms and viruses were sampled by peristaltic pump followed by on-deck filtration. Several filtration steps were performed using membranes with different pore sizes to obtain size-fractionated samples corresponding to viruses (<0.1 µm and 0.1-0.2 µm), prokaryotes (0.2-3 µm) and eukaryotes (0.8-5, 5-20, 20-180 and 180-2000 µm). In this study, we only used samples collected from the surface water layer. Details about genomics methods are available in (Carradec et al, 2018) and in the following publications: Virus metagenomes (Roux et al., 2016); Prokaryote metagenomes (Sunagawa et al., 2015); Eukaryote metabarcoding (de Vargas et al., 2015); Eukaryote metagenomes and metatranscriptomes (Alberti et al., 2017; Carradec et al., 2018). The abundance of individual genes was assessed by normalization to the total number of sequences within the same organismal group (Carradec et al, 2018). Cyanobacterial clade absolute cell abundance was assessed using the *petB* marker gene, as described in (Farrant et al., 2016), in combination with flow cytometry counts using the method published by (Vandeputte et al., 2017).

Metatranscriptomic and metagenomic unigenes were functionally annotated using PFAM (Finn et al., 2016) as the reference database and search tool (Durbin et al., 1998). To detect the presence of genes encoding silicon transporters, ferritin, proteorhodopsin, FBAI and FBAII among the unigene collection, the profile hidden Markov models of the PFAMs PF03842, PF0210, PF01036, PF00274 and PF01116, respectively, were used, with HMMer v3.2.1 with gathering threshold option (http://hmmer.org/). It is important to note that flavodoxin (PF00253,

PF12641 and PF12724), ferredoxin (PF00111), and cytochrome c_6 (PF13442) PFAM families do not discriminate those sequences involved in photosynthetic metabolism from other homologous sequences. The photosynthetic isoforms for flavodoxin, ferredoxin and cytochrome c_6 were therefore determined by phylogeny, as described below.

To discriminate the photosynthetic isoforms from other homologous sequences, we started with the results from HMMer and then built libraries composed of well-known reference sequences (manually and experimentally curated) from both photosynthetic and non-photosynthetic pathways. To enrich our libraries we used the reference sequences to find similar sequences by using BLAST search tool ("tBLASTn" program with an e−5e value threshold) against phyloDB reference database (Dupont et al., 2015). Next, we used MAFFT version 7 using the G-INS-I strategy (Katoh and Standley 2013). The corresponding phylogenetic reference trees were generated with PhyML 3.0 (Guindon et al. 2010) using the LG substitution model with four categories of rate variation. The starting tree was a BIONJ tree and the type of tree improvement was subtree pruning and regrafting. Branch support was calculated using the approximate likelihood ratio test (aLRT) with a Shimodaira–Hasegawa-like (SH-like) procedure. We then manually identified the branches containing the photosynthetic versions and those with non-photosynthetic proteins. We ensured that the aLRT values of the photosynthetic and non-photosynthetic branches were higher than 0. 7 by retaining only the most conserved matches in our trees. Finally, we realigned and labeled the unigenes against the reference trees depending on the placement of each translated unigene on them.

While HMMer has the highest sensitivity among the classical domain detection approaches, not all the references collected by PFAM are sufficiently rich with HMMer to maintain the same detection (Bernardes et al., 2016). To deal with the poor representation of *ISIP* genes in the PFAM database and to improve their detection, we adopted a simplified version of the approach presented in (Bernardes et al., 2015) to build our own pHMM to detect the different conserved regions represented by ISIP1, ISIP2a and ISIP3 amino acid sequences. For this, we collected all the sequences in the reference literature (Allen et al., 2008; Chappell et al., 2015; Morrissey et al., 2015; Lommer et al., 2015), all 35 sequences belonging to PFAM PF07692, and the 56 most conserved sequences from PF03713 (all the seeds).

2.8. Data and code availability

Sequencing data are archived at ENA under the accession number PRJEB4352 for the metagenomics data and PRJEB6609 for the metatranscriptomics data (Carradec et al., 2018). Environmental data are available at PANGAEA. The gene catalog, unigene functional and taxonomic annotations, and unigene abundances and expression levels are accessible at http://www. genoscope.cns.fr/tara/. Computer codes are available upon request to the corresponding authors.

2.8.1 Accession numbers of metagenomics and metatranscriptomics data

2.8.1.1 Sample:

ERS492651, ERS492651, ERS492650, ERS492669, ERS492669, ERS492662, ERS492662, ERS492658, ERS492658, ERS492650, ERS492740, ERS492742, ERS492742, ERS492751, ERS492751, ERS492763, ERS492763, ERS492757, ERS492740, ERS492757, ERS492825, ERS492825, ERS492824, ERS492824, ERS492829, ERS492852, ERS492852, ERS492846, ERS492846, ERS492829, ERS492897, ERS492897, ERS492895, ERS492895, ERS492912, ERS492912, ERS492909, ERS492909, ERS492904, ERS492904

2.8.1.2 Experiment:

ERX948080, ERX948010, ERX1782415, ERX1782384, ERX1782327, ERX1796912, ERX1796638, ERX1796690, ERX1796805, ERX1782126, ERX1782109, ERX1782245, ERX1796854, ERX1796544, ERX1782292, ERX1782172, ERX1782221, ERX1796700, ERX1796855, ERX1782301, ERX1782464, ERX1782128, ERX1789668, ERX1789366, ERX948029, ERX948074, ERX1796627, ERX1796773, ERX1789369, ERX1789449, ERX1796931, ERX1796605, ERX1789426, ERX1789575, ERX1789524, ERX1796866, ERX1796524, ERX1789649, ERX1789612, ERX1789647, ERX1796596, ERX1796836, ERX1789655, ERX1789574, ERX1789407, ERX1782118, ERX1782283, ERX947973, ERX948088, ERX1789391, ERX1789539, ERX1789587, ERX1796687, ERX1796586, ERX1796703, ERX1789662, ERX1789616, ERX1789589, ERX1796662, ERX1796518, ERX1796678, ERX1796698, ERX1782217, ERX1782352, ERX1796645, ERX1796858, ERX1796924, ERX1789675, ERX1789597, ERX1789700, ERX1789362, ERX1782350,

ERX1782418, ERX947994, ERX948064, ERX1789361, ERX1789368, ERX1789532, ERX1796658, ERX1796818, ERX1796632, ERX1789638, ERX1789548, ERX1789579, ERX1796921, ERX1796732, ERX1796741, ERX1789714, ERX1789489, ERX1789628, ERX1796689, ERX1796850, ERX1796523, ERX1782181, ERX1782370, ERX1796607, ERX1796738, ERX1796714, ERX1789437, ERX1789516, ERX1789417

2.8.1.3 Run:

ERR868475, ERR868513, ERR1712182, ERR1712118, ERR1711869, ERR1726556, ERR1726667, ERR1726938, ERR1726688, ERR1712207, ERR1711933, ERR1711897, ERR1726927, ERR1726932, ERR1712069, ERR1712197, ERR1711986, ERR1726883, ERR1726891, ERR1712219, ERR1711929, ERR1711951, ERR1719463, ERR1719159, ERR868466, ERR868469, ERR1726762, ERR1726913, ERR1719393, ERR1719310, ERR1726961, ERR1726522, ERR1719437, ERR1719413, ERR1719343, ERR1726622, ERR1726721, ERR1719297, ERR1719410, ERR1719307, ERR1726770, ERR1726561, ERR1719256, ERR1719298, ERR1719217, ERR1711914, ERR1711917, ERR868363, ERR868489, ERR1719301, ERR1719160, ERR1719214, ERR1726564, ERR1726725, ERR1726569, ERR1719448, ERR1719389, ERR1719194, ERR1726571, ERR1726533, ERR1726892, ERR1726601, ERR1711949, ERR1712155, ERR1726608, ERR1726657, ERR1726763, ERR1719391, ERR1719175, ERR1719381, ERR1719365, ERR1711882, ERR1711999, ERR868382, ERR868352, ERR1719395, ERR1719316, ERR1719207, ERR1726643, ERR1726714, ERR1726846, ERR1719404, ERR1719213, ERR1719459, ERR1726822, ERR1726912, ERR1726691, ERR1719356, ERR1719145, ERR1719293, ERR1726695, ERR1726666, ERR1726903, ERR1712102, ERR1711923, ERR1726745, ERR1726946, ERR1726765, ERR1719295, ERR1719249, ERR1719385

3. Results

3.1 Modeled iron distributions are highly correlated with the expression of marker genes for iron limitation

Iron is a complex contamination-prone micronutrient whose bioavailability is difficult to assess in the ocean (Tagliabue et al., 2017). Rather than using single discrete measurements, we linked observed differences in plankton communities at sites sampled during the *Tara* Oceans expedition (Bork et al., 2015) with the range of iron conditions typical of each location.

Specifically, we extracted annual mean iron concentrations and their variability from two state-of-the-art ocean models (ECCO2-DARWIN (Menemenlis et al., 2008) and PISCES2 (Aumont et al., 2015)), and analyzed their correspondence with the best available estimates based upon *in situ* data (a compilation of iron observations (Tagliabue et al., 2012) merged with GEOTRACES data (Tagliabue et al., 2012; Mawji et al., 2014) in a manner similar to previous studies (Toulza et al., 2012) (Figure 1).

To assess the reliability of the modeled iron distributions, we correlated the expression of diatom *ISIP* genes in metatranscriptomics datasets with the annual means of iron concentrations estimated by the DARWIN model, and with annual and monthly means by the PISCES2 model (Carradec et al., 2018) (Table S1a and Supplementary Material S1). These genes have been found in multiple previous studies to be inversely correlated with iron availability (Allen et al., 2008; Chappell et al. 2015; Morrissey et al., 2015, Graff van Creveld et al., 2016; Marchetti et al., 2017). Figure 1 presents a comparison between the estimates of dissolved iron concentrations derived from the annual mean iron field from PISCES2 and ECCO2-DARWIN (Table S1a), with the *Tara* Oceans stations superimposed and best available estimates based upon *in situ* measurements (Figure 1). In spite of the evident scarcity of actual iron concentration data (which illustrates the need to use models for estimating iron in the current exercise; Figure 1), both models and *ISIP* mRNA levels describe very satisfactorily the global scale gradients, with the highest concentrations of iron observed in the Mediterranean and Arabian Seas (both highly impacted by desert dust deposition) and the lowest in the tropical Pacific and Southern Oceans. This demonstrates that the geographical coverage of the *Tara* Oceans expedition is well suited to studies of the role of iron on euphotic planktonic ecosystems. The available data (Figure 1, Table S1a) further indicates that the gradients of iron appear to be better captured by PISCES2, a more complex and recent model (Aumont et al., 2015). This is for instance the case for the North Atlantic Ocean and the Mediterranean Sea, where longitudinal gradients are stronger in PISCES2 and are consistent with *ISIP* gene levels, while ECCO2-DARWIN seems to overestimate iron in the Eastern Atlantic Ocean and underestimate it in the Mediterranean Sea. The opposite is true in the South Atlantic Ocean, where *ISIP* mRNA levels show a clear increase correlated with iron stress between South America and Africa (Figure 1). Overall, in the Atlantic ECCO2-DARWIN has higher concentrations, and thus a clearer large scale Atlantic-Pacific gradient is observed.

The Pacific and Southern Oceans (subpolar and polar stations TARA 81-85) are both characterized by low levels of iron, as mentioned above. Notably, PISCES2 has a rather flat distribution in the Pacific Ocean, with very low values, while the other model shows a relatively higher level of iron at the core of the subtropical gyres, i.e., close to the Hawaii Islands (Stations TARA_131 and TARA_132) and offshore from South America (TARA_98 and close-by stations) that seems to be in agreement with *ISIP* mRNA levels (at least for the Hawaiian sample – Figure 1). These are very oligotrophic oceanic regions, where nitrate is also a strongly limiting nutrient. Again, the *ISIP* expression pattern in Figure 1 is closer to the PISCES2 model, in that it shows a clear reduction of the stress resulting from iron deprivation within these gyres. Finally, while a significant increase in iron at the Equator may be expected as a consequence of the upwelling in this region, both the models and the *ISIP* levels (at Station TARA_128) suggest that this area is rather characterized by low values of iron. Overall, our analysis indicates that both models correlated very well with *Tara* Oceans transcriptomic data, with no relevant differences among monthly and yearly values, but with annual means from the ECCO2-DARWIN estimates showing the best reliability (Table S1c). This analysis also indicates that metatranscriptomics is now mature enough to provide an independent, biologically-based validation of ecosystem models.

3.2 Plankton response to iron availability is coordinated at sub-community level

The higher level organization of plankton communities, and its possible relationship with the roles of individual constituents, has been highlighted previously in an analysis of the potential links between community structure and carbon export using data from *Tara* Oceans (Guidi et al., 2016). We here used this approach to explore plankton ecosystem responses to iron bioavailability using an end-to-end approach from genes to communities and from viruses to metazoa to reveal community responses at global scale (see Methods). Known as weighted gene correlation network analysis (WGCNA; see Methods for further description, Guidi et al., 2016; Langfelder and Horvath, 2007), this approach deciphers sub-communities (modules) of organisms within a global co-occurrence network, and because of the high levels of co-variation of individual taxa it is possible to deduce putative ecological interactions. As proxies for organism abundance we used the relative abundances of eukaryotic lineages (defined as operational taxonomic units; OTUs) derived from 18S-V9 rDNA metabarcoding data (de Vargas

et al., 2015). WGCNA generated a total of 31 modules. Each module groups a subset of eukaryotic taxa found in *Tara* Oceans samples whose pairwise relative abundance was highly correlated over all the sampling sites, i.e., they have a high probability of co-occurrence and to change their abundance in a coordinated way. Because they react in phase, we can infer that within each sub-community these organisms have a higher probability of interaction among themselves than with the organisms in other modules.

We found four eukaryotic subnetworks significantly associated with the ECCO2-DARWIN-derived and/or with the PISCES2-derived estimates of iron concentrations in the global ocean (Figure 2a, Figure S1a, b, Table S1d). The Black and Turquoise modules were associated with high significance to the iron concentrations generated by both models whereas the DarkRed and Yellow modules were better associated with ECCO2-DARWIN and PISCES2, respectively: Black (DARWIN: R=0.37, $P=6x10^{-4}$; PISCES2: R=0.38, $P=3x10^{-4}$), Turquoise (DARWIN: R=0.46, $P= 1x10^{-5}$; PISCES2: R=0.42, $P=9x10^{-5}$), DarkRed (DARWIN: R=-0.43, $P=5x10^{-5}$; PISCES2: R=0. 19, P=0.08), and Yellow (DARWIN: R=0.19, P=0.09; PISCES2: R=0.56, $P=5x10^{-8}$), and contained between 31 and 591 different OTUs (Tables S1d and S1e). These subnetworks were denoted iron-associated assemblages (IAAs). For each IAA subnetwork, WGCNA computes a single representative as a combination of lineages. Such a score, denoted a 'module eigengene' score (hereafter termed an eigenlineage score) represents the first eigenvector of the assemblage (Langfelder and Horvath, 2007). Projections of samples on such an eigenvector show the relative importance of samples to the global variance of each IAA. Together with their contribution, in terms of OTU abundance to the total eukaryotic abundance in each station (Table S1f), they provide clues to interpret the link between modules and iron availability. The mismatch in some regions between the two models (see above) is likely the reason why the significance of association of the Yellow module with ECCO2-DARWIN, whose variance and representativeness is particularly significant in the South Adriatic and is minimally present in the Peruvian upwelling area, is much less than that with PISCES2. By contrast, the DarkRed module, which appears to be the best indicator module for the Marquesas area (Figure 2b, upper panel) and is highly relevant in the Peruvian upwelling region, displays a much less significant association and an opposite variation with PISCES2 iron vs. ECCO2-DARWIN iron. The IAAs show slightly different, often antagonistic, variance contributions at global scale

(Figure 2b, upper panel), with each of them being particularly responsive, in terms of variance, in specific sites, e. g., the Yellow module in the Eastern Mediterranean Sea.

We examined the lineage composition of each IAA and the relevance of each taxon within them by determining the relative abundance of each lineage with respect to iron concentration estimates and their centrality within the module (see Methods). The results are reported in Table S1d,e. The IAAs displayed significant differences in terms of numbers of lineages and compositions, with the Turquoise module being the largest and dominated by consumers, predominantly metazoans, and the DarkRed module being the smallest. The Black module displayed the highest proportion of autotrophs, while the DarkRed IAA displayed the highest proportion of diatoms (Bacillariophyta; 57% of all autotrophic protists).

To reduce complexity further, we screened the networks in terms of the VIP score of each node (i.e., the OTUs displaying the highest statistical weight in differentiating sites because of iron availability; Methods, Table S1d, Figure 2a and Figure S1c). Species with high VIP scores can be predicted to be particularly important in reflecting the adjustments of each module via their specific interactions with other members of their sub-community. Although interpreting why high VIP taxa are related to iron bioavailability is often severely restricted by our knowledge of plankton functional ecology and inter-organism interactions, in other cases the role of VIP taxa within the modules is clearer. As an example, identification of an IAA in which several diatoms have the highest VIPs (DarkRed module, 8 subnetwork members, -0.337 correlation with iron), commonly found in the most severely iron-limited regions of the world's ocean and often the most responsive groups in mesoscale iron fertilization experiments (Boyd et al., 2007; Marchetti et al., 2006), suggests a strong physiological plasticity of these groups (Greene et al., 1991; Lommer et al., 2012). The fact that *Pseudo-nitzschia* is among the highest scoring VIP genera in the DarkRed module further suggests that this genus tracks regions with low iron bioavailability, being able to profit from it when it becomes available. Other examples concern metazoans: copepods from the genus *Temora* (high subnetwork centrality, strong correlation with iron) are known to be iron-limited (Chen et al., 2011), and the two cnidarian lineages - the class *Hydrozoa* and the genus *Pelagia* (both of which display relatively strong subnetwork centrality and strong correlations with iron) - suggest strong predator-prey links.

Considering the ECCO2-DARWIN-derived VIP scores, lineages with the highest scores (>1) could predict as much as 61.9%, 52.6%, 49.1%, and 38.1% (in the Turquoise, Black, DarkRed and Yellow IAAs, respectively; leave-one-out cross validated - LOOCV) of the variability of iron in the oligotrophic ocean. When the PISCES2-derived VIP scores are taken into account, the predictive potential of the IAAs is even higher: 73.2% (Turquoise), 61.9% (Yellow), 59.0% (Black) and 54.4% (DarkRed). More importantly, the VIP scores obtained with the two models for each OTU showed an extremely good covariance (Figure S1d). This confirms the biological coherence and stability of the modules and their components to iron availability despite the occasional mismatch in the predictions of the two models.

Of the photosynthetic groups, autotrophic dinoflagellate taxa were particularly relevant in the Turquoise and Black modules, diatoms were relevant in the DarkRed module, and haptophytes were significantly present in the Yellow module. Metazoans were particularly important in the Black and the Turquoise modules, and marine stramenoplies (MAST)/marine alveolata (MALV) groups of phagotrophic and parasitic heterotrophs were relevant in the Black (MALV), Turquoise and Yellow modules (MAST) (Figure 2a, Figure S1c, and Table S1d, e). This hints at particularly intricate, and still elusive, interactions among organisms which ultimately lead to the observed collective responses.

To further interpret the patterns observed for the IAAs we chose two additional modules, denoted DarkGrey and Red, because of the different correlations of diatoms within these modules to iron concentrations with respect to the DarkRed module (Figure S1a, c). By examining the abundance of the components of each module at different sampling sites (Table S1f), the results suggest that the Turquoise module groups lineages relevant in all of the main oceanic biogeographic regions with the exception of the Mediterranean basin, and with a prominent weight in the Southern Ocean. By contrast, the Black and Yellow modules are of particular importance in the Mediterranean Sea, while other IAAs have minor contributions. The DarkRed module is generally poorly represented, however in the South Pacific and in particular around the Marquesas Islands, its relevance is high (Figure 2b, upper panel and Table S1f).

Based on all of the above information, we then sketched the ecological profiles of the seven modules, summarized below:

573 Black IAA: Ubiquitous, but with low abundance except in the Mediterranean basin, and
574 composed principally of heterotrophic organisms (protists and metazoans) (Table S1e, f).
575 Dinophytes are the autotrophic component of this module while diatoms are poorly represented.
576 Around the Marquesas Islands, its weight is constantly low. Lineages are positively correlated or
577 loosely anticorrelated with iron (Figure 2a, b and Table S1d). This module has an intermediate
578 level of internal connectivity and suggests top heavy (pyramidal) trophic interactions. The
579 assemblage resembles a typical pattern in a post-bloom phase, with biomass accumulated in the
580 metazoan compartment. No significant differences are seen when the ECCO2-DARWIN-derived
581 and PISCES2-derived VIPs are compared since the module is not relevant in areas where the two
582 models disagree. This pattern is consistent with the differences detected at molecular level.

583 DarkRed IAA: The module is not particularly significant at global scale in terms of abundance
584 (Figure 2b, upper panel and Table S1e). It contains a small number of lineages with a high
585 relative weight of diatoms and few metazoans but no copepods, with carbon recycling mostly in
586 the protistan compartment. This module is particularly intriguing because, with very few
587 exceptions, all the lineages including diatoms are negatively correlated with iron (Figure S1c). It
588 is particularly responsive in the Marquesas area but is also present in offshore South American
589 upwelling areas. The internal connectivity is of an intermediate level (Table S1d). These features
590 hint at an assemblage in the subtropical ocean driven by the activity of diatoms thriving in
591 regions of low iron availability (while exploiting a higher than average silicon availability), thus
592 showing an inversion of the pattern compared to high iron regions (Figure 2b, upper panel).
593 Significantly, its abundance drops at Station TARA_123 in the Marquesas archipelago (see
594 below).

595 Turquoise IAA: Ubiquitous, with a general high weight in terms of abundance, and very
596 abundant in the Southern Ocean (in particular in stations TARA_85-88; Table S1f). The module
597 includes relatively few diatoms, but many dinoflagellates (both autotrophic and heterotrophic
598 species) (Tables S1d, e). Copepods are the most numerous components and show the highest VIP
599 scores. Of note, this module includes the crustacean order *Euphausiacea* (krill), which
600 specifically emerges as having high VIP scores only when the PISCES2-derived iron estimates
601 are used. Both internal connectivity and number of lineages are high (Table S1d). The module as

a whole responds in the Marquesas area, especially at TARA_123 (Figure 2b, upper panel and Table S1f).

Yellow IAA: This module is particularly important in South Adriatic and Eastern Mediterranean, as well as in the tropical North Atlantic (Figure 2b, upper panel and Table S1f). It includes relatively few metazoans and diatoms but a notable abundance of haptophytes and heterotrophic protists (Table S1d, e). Displays a weak response in the Marquesas area (TARA_125) (Figure 2b, upper panel) and seems to be less dependent on iron availability as compared to the other modules.

DarkGrey: Not an IAA and has a low weight in general, with a slight positive correlation to iron and only low internal connectivity. Diatoms in this module are very relevant (Table S1d, e). Contains a high fraction of metazoans with fewer heterotrophic protists. This module displays a typical bottom heavy (pyramidal) structure with diatoms reacting positively to iron availability.

Red module: Not an IAA, but this module displays a similar response to iron than the DarkGrey module, with the main differences being that it contains few metazoans and the protist compartment is dominated by Dinophyceae. Diatoms are also dominant as autotrophic protists. It is the module that correlates the most with chlorophyll and primary productivity (Figure S1a) and seems to be associated with highly productive areas. It is thus not very relevant globally, with the exception of the South Atlantic Ocean, where it dominates the Benguela upwelling (Station TARA_67), a very rich region that is not iron limited. It is apparently driven by bottom-up flexible responses to iron availability, most probably by macronutrient availability (Table S1d, e). It displays variable correlations of its members to iron, and has also a bottom heavy pyramidal trophic structure.

Overall, our analysis strongly suggests that different sub-assemblages of co-occurring lineages can be pinpointed within communities that respond differently to resource limitation, mostly without marked geographical preferences albeit with high plasticity to iron availability. Particularly remarkable is the contrasting role shown by diatoms, with different lineages covering the full range of correlations with iron (Figure S1c), possibly linked to their different strategies for responding to the lack of a crucial resource. In some cases their communities share a similar response while in others the structure of the assemblage is modified. The further observation that co-occurrence of IAAs can show biogeographical patterns (Figure 2b, upper

panel) that are not clearly emphasized by analysis of single eukaryotic groups (Figure S1c) is suggestive of a compartmentalization of communities in sub-communities or modules. Our analysis also infers that it is the module as a whole that responds to perturbation, reinforcing the need to dissect plankton responses to iron bioavailability at community scale, while investigating the physiological responses of key species.

In addition to eukaryotes, WGCNA analysis was also performed on prokaryotic communities, as well as on prokaryotic genes from the Ocean Microbial Reference Gene Catalog (Sunagawa et al., 2015; Alberti et al., 2017). Using relative abundances of prokaryotic 16S rDNA miTags, no subnetwork could be associated significantly with iron (maximum r=0.19, $P<10^{-2}$). However, following the same procedure but using the relative abundances of prokaryotic genes rather than taxa, five subnetworks were significantly associated with iron (ECCO2-DARWIN iron data; Figure 2b, lower panel and Table S1g) ($P<10^{-5}$): Grey60 (r= 0.38, P=6. 10^{-5}), Plum1 (r= 0.54, P=3. 10^{-9}), Red (r=-0.42, $P=10^{-5}$), SkyBlue (r=-0.44, P=2. 10^{-6}), and SaddleBrown (r=-0.47, P=6. 10^{-7}). VIPs obtained from each of the two models displayed high correlations (Grey60= 0.99, Plum1=0.94, Red=0.99, SkyBlue=0.96, SaddleBrown=0.98). The VIP genes of the SaddleBrown subnetwork represent 25% (N=41) of the total number of genes, and several genes that could be functionally identified encode proteins associated with iron transport, saccharopine dehydrogenase, aminopeptidase N, and ABC-type transporters (Table S1g). The Plum1 subnetwork is a small subnetwork of around 100 genes that is solely associated with iron concentration variability, and 30% of its VIP genes encode principally specialized functions defined as non-core functions in a previous study of the *Tara* Oceans Global Ocean Microbiome (Sunagawa et al., 2015) (Table S1g). Not surprisingly, 75% of the genes within this subnetwork encode proteins with unknown functions, although some known functions are linked to iron, such as ferredoxin and regulation of citrate/malate metabolism. The contribution to the global variance by stations located within the Red Sea (Stations TARA_31-34) is particularly high (Figure 2b, lower panel). The Red subnetwork is very large, composed of 3,059 genes. However, only 9% represent high scoring VIPs, among which functions related to iron are evident (e.g., ABC-type Fe^{3+} siderophore transport system, putative heme iron utilization protein, metalloendopeptidase - Table S1g). Finally, the SkyBlue subnetwork is a small subnetwork (172 genes) containing 33% of VIPs whose functions are generally unknown (Table S1g). The global

variance of this gene subnetwork can be correlated principally with several oligotrophic regions of the Pacific Ocean (e. g., Stations TARA_93, 100, 112, 128).

In summary, association of prokaryotes with iron is detectable at the functional level (gene abundance) but not at the taxonomic level, which would suggest a low level of specialization, at least with the resolution allowed by the 16S marker. To further analyze this aspect we focused on *Prochlorococcus* and *Synechococcus*, the two most abundant and widespread bacteriophytoplankton in the global ocean, and for which a higher resolution genetic marker is available. Combining the information from the taxonomic marker *petB*, which encodes cytochrome b_6 (Farrant et al., 2016), with flow cytometry cell counts we estimated the absolute cell abundance of the picocyanobacterial clades and found that many of them have a strong correlation with predicted iron levels from PISCES2 (Figure 3) and ECCO2-DARWIN models (not shown). *Prochlorococcus* HLIII and IV ecotypes showed the highest anti-correlation with iron, in agreement with previous descriptions that they are the dominant populations in HNLC areas (Rusch et al., 2010; West et al., 2011). Prochlorococcus LLI, a minor component in surface waters, also showed anticorrelation with iron. In the case of *Synechococcus*, the strongest positive correlation was found for clade III, whereas a weaker pattern is displayed by clade II. On the contrary, CRD1 showed the highest negative correlation with iron, consistent with it being reported as the major *Synechococcus* clade in HNLC regions (Farrant et al., 2016; Sohm et al., 2016). In addition, clade EnvB also displayed a negative correlation with estimated iron concentrations.

These results demonstrate that iron affects picocyanobacterial community composition and raise the question of whether the lack of correlation with taxonomic networks depends on a poor taxonomic resolution or to being more pronounced for autotrophs with respect to heterotrophs.

Finally, we used relative abundance of viral populations (Brum et al., 2015) to apply WGCNA and tentatively explore whether the viral module subnetworks display any kind of association to the same suite of environmental factors used above for prokaryotes and eukaryotes (data not shown). In spite of the fact that we found four viral IAAs significantly associated with iron using the ECCO2-DARWIN iron estimates (data not shown), our current knowledge of marine viruses is not advanced enough to discuss our results in the view of the impact on viruses of global iron biogeochemistry. This lack of knowledge is aggravated by the fact that the vast majority of viruses in the IAAs have unknown host ranges.

693 Viruses are thought to impact oceanic iron during host lysis; however, there is a current
694 discussion about their potential role in complexing iron (Bonnain et al., 2016). To explore this
695 latter point, we surveyed the *Tara* Oceans metagenomes for genes encoding viral structural
696 proteins with putative iron-binding sites. Specifically, we searched for paired histidine residues
697 (HxH motifs) in tail proteins (Bartual et al., 2010) and baseplate assembly proteins (Browning et
698 al., 2012) because this motif has been experimentally implicated in the octahedral coordination
699 of iron. We also analysed the presence of four conserved cysteine residues involved in the
700 coordination of a 4Fe-4S cluster in tail tip proteins (Tam et al., 2013). Remarkably, these
701 potential iron-binding motifs are present in 87% unigenes encoding viral tail proteins, 47% of
702 baseplate assembly proteins, and 12% in those coding for tip proteins (Figure 4a-c). The
703 corresponding viral contigs are distributed ubiquitously and with high abundance (Figure 4a-c),
704 suggesting that a significant fraction of colloidal iron may be associated with viruses in the
705 ocean, a factor that is not currently considered in the modeling of ocean biogeochemistry. The
706 question is then how substantial this contribution could be. (Bonnain et al., 2016) made a broad
707 estimation based on the number of iron ions experimentally determined in tails of non-marine
708 phages, and the amount of tailed viruses typically found in marine surface waters. They thereby
709 suggested that between 6 and 70% of the colloidal iron fraction from surface waters could be
710 bound to tail fibers of phages. In this context, the recent "Ferrojan Horse Hypothesis" posits that
711 iron ions present in phage tails enable phages to exploit their bacterial host's iron-uptake
712 mechanism, where the apparent gift of iron leads to cell lysis (Bonnain et al., 2016). Although
713 our analysis does not allow to confirm this hypothesis, it provides a useful context to explore it
714 further.
715

716 **3.3 Functional responses are mediated either by changes in gene copy number or by**
717 **expression regulation**

718 Given the clear patterns in the community responses to iron availability, we next wondered
719 which molecular patterns were associated with them. We first examined the prevalence of the
720 diatom *ISIP* genes in more detail using both metagenomics and metatranscriptomics data to
721 detect changes in gene abundance and expression, respectively. We found that both the
722 abundance and expression of this gene family displayed a strong negative correlation with iron
723 (Figure 5a). Figure 5a shows a strong hyperbolic profile of *ISIP* gene abundance and mRNA

levels with respect to iron concentrations (non-linear regression fitness of 97.01 and 98.14, respectively; Table S1c). Furthermore, density clustering algorithms detected two types of responses – stations in which *ISIP* was only increased in metagenomics data (denoted Group 0) and others in which both metagenomic and metatranscriptomic data showed increases in *ISIP* levels (denoted Group 1; Figure 5a). The former likely correspond to locales where *ISIP* copy numbers vary in diatom genomes as a function of iron, implying that the diatoms at these stations display permanent genetic adaptations to the ambient iron concentrations, whereas the latter display transcriptional variation, indicative of more flexible short-term acclimatory rather than permanent adaptive evolutionary processes. Taxonomic analyses revealed that diatoms from the *Thalassiosira* genus were typical of Group 0, whereas *Pseudo-nitzschia* was found largely in Group 1 (Figure 5b). Representatives from both these genera are well known to respond to fluctuations in iron (Cohen et al., 2017) see Supplementary Material S1), so these different iron-response strategies may underlie why they are present in different IAAs; *Thalassiosira* is present in the Black and Turquoise IAAs whereas *Pseudo-nitzschia* is only present in the DarkRed module, where it is negatively correlated to iron (Figure 2a, Table S1d).

It is interesting to note that sampling sites can be grouped in a similar way according to either their picocyanobacterial community or diatom *ISIP* patterns in relation to iron levels (Figure 6). HLIV and HLIII codominate the *Prochlorococcus* community in Group 1 stations, and these sites are also characterized by the presence of LLI, as well as the *Synechococcus* clades CRD1 and EnvB. Based on picocyanobacteria community composition, these stations tend to cluster together in a group of low-iron stations from Indian and Pacific Oceans (TARA_52, 100, 102, 110, 111, 122, 124, 125, 128, 137). On the contrary, Group 0 ISIP stations were dominated by either *Prochlorococcus* HLI or HLII and by *Synechococcus* clades II or III. Among these stations, those from the high-iron Mediterranean Sea (TARA_7, 9, 18, 23, 25, 30) clustered together based on picocyanobacteria community composition (Figure 6).

Besides diatoms, we carried out a detailed analysis of *ISIP* distributions among other phytoplankton taxa. We found that in chlorophytes Fea1-domain-encoding genes (related to *ISPI2a*; Marchetti et al., 2017) vary in copy number as a function of predicted iron levels, and that *ISIP* expression also varies in haptophytes and pelagophytes (Figure S2). Dinoflagellates display the lowest correlations of *ISIP* gene abundance and expression with respect to iron. This

may indicate that dinoflagellates respond differently to iron concentrations or with different genes.

A similar analysis was performed to examine the abundance and expression of type I (metal-free) and type II (containing iron or other divalent cations) FBAs (Allen et al., 2012). We found that the *FBAII* gene showed a clear up-regulation at high-iron stations in all groups, while diatoms showed a concomitant reduction in *FBAI* gene abundance and mRNA levels, pelagophytes and dinoflagellates displayed decreased gene abundance, while haptophytes displayed a response at the mRNA level (Figure S2). The chlorophytes displayed no consistent trends.

Ferritin is another important protein of iron metabolism that was relatively recently identified in diatoms (Marchetti et al., 2009) and in other phytoplankton functional groups (Botebol et al., 2015). Although it appears to be involved in long-term iron storage in *Pseudo-nitzschia* (Marchetti et al., 2009), other studies have suggested that its principal role could be in cellular iron buffering and temporal storage over shorter timescales such as during diurnal cycles (Pfaffen et al., 2015; Botebol et al., 2015; Cohen et al., 2018). Our analysis revealed no clear pattern in ferritin gene abundance or expression and estimated iron levels (Figure S2), suggesting that iron storage may not be the main function of ferritin in most eukaryotic marine phytoplankton. The exceptions are haptophytes, in which an iron-driven increase in copy number is observed (Figure S2), and the diatom genus *Pseudo-nitzschia*, in which the biogeographical patterns of ferritin gene abundance and expression suggest a positive correlation with iron (Figure S3b).

We additionally examined the levels of genes encoding proteorhodopsin, a light-driven proton pump for the generation of ATP that has been proposed to supplement ATP generation from photosynthesis in iron-limiting conditions, when photosynthetic electron transport is sub-optimal (Marchetti et al., 2015). According to our results, abundance of the gene is negatively correlated with iron in pelagophytes and dinoflagellates (as well as in diatoms, albeit without statistical support), and mRNA levels are negatively correlated with iron availability in pelagophytes and haptophytes (Figures S2).

We also examined the interaction between iron and other nutrients in diatoms. Particularly, we focused on silicate metabolism because iron bioavailability has been found to play a role in silicon utilization in these organisms (Mock et al., 2008; Durkin et al., 2012). The analysis of the

different clades of Si transporter (SIT) multigene family support the strong interaction between iron and silicate in diatoms and suggest that the diversification of SITs has led to specialized adaptations to deal with it (see Supplementary Material S1).

Collectively, our results indicate that individual genes implicated in iron metabolism in specific organismal groups do not provide an unequivocal evaluation of iron availability in the environment, and are thus of only limited use as sentinel genes of iron bioavailability. Instead, the integration of all these iron-driven patterns, spanning from genes to ecosystems, is a promising strategy for designing omics-enabled tools that can improve the representation of key nutrients in biogeochemical models. In this sense, the co-variation of picocyanobacterial communities with the transcriptional regulation and altered copy numbers of diatom *ISIP* genes can potentially be exploited to predict actual iron bioavailability in the ecosystem (Figure 6). A recent report on the phytoplankton transcriptional response to upwelling (Lampe et al, 2018) highlighted that diatoms express genes involved in nitrogen assimilation, probably to overcome possible autotrophic competitors, thus suggesting that different transcriptional sets of genes may be expressed under different bloom-triggering conditions.

3.4 Plankton respond to a resource burst in the Marquesas archipelago by reorganization of IAAs

The global analyses of IAAs and iron responsive genes in the context of the ranges of geographic iron availability provide a first order approximation of plankton community structure organization and responses for large scale, iron-linked biogeochemical regions. In other words, they possibly reflect the integrated, albeit diversified, response to average conditions and in a stationary or quasi-stationary phase. They further provide support for the iron products of the two biogeochemical models. We reasoned that they might also be able to indicate increases in iron in regions where biogeochemical models do not have sufficient resolution and to highlight mechanisms in action when the resource is provided in bursts that drive the community out of a previous steady state, e.g., leading to blooms. One such case is the Marquesas archipelago in the sub-tropical Pacific Ocean, where previous studies (Martinez and Maamaatuaiahutapu, 2004) have highlighted a dynamic natural perturbation resulting in perennial plankton blooms that are visible from space. Although iron concentrations have not been measured extensively in the region, these and similar blooms (Gong et al., 2016) are triggered by different processes due to

814 the presence of the islands (vertical mixing, horizontal stirring, local precipitation and runoff)
815 which are typically coupled to iron injection (Martinez and Maamaatuaiahutapu, 2004), a
816 phenomenon that has been termed Island Mass Effect (Gove et al., 2016). We therefore focused
817 on this region to examine the relationship between the global patterns in plankton sub-
818 communities and iron-responsive gene abundance/expression in a more localized dynamic
819 setting (Supplementary Material S2-S4).
820 Satellite chlorophyll estimates showed that in the days preceding the visit of the *Tara* Oceans
821 expedition to the archipelago in August 2011 the area was characterized by intense variability.
822 Our analyses also revealed a highly turbulent environment, with mixing up to 100 m depth and
823 strong lateral shearing downstream of the islands, which generated an area of recirculation in the
824 wake of the main island and the formation of small eddies where the blooms were occurring
825 (Figs. 7a and S4a). Station TARA_122 sampled the HNLC pre-bloom waters upstream of the
826 islands (Figure 7a). Waters of Station TARA_122 were characterized by low chlorophyll
827 concentrations in the water column ($[\text{Chl-}a]_{\text{int:}}$ 16.6 $\text{mg}\cdot\text{m}^{-2}$) but high concentrations of nutrients
828 (NO_{2-}: 0.12 mmol. m^{-3}, PO_4: 0.57 mmol. m^{-3}, NO_2NO_3: 5.5 mmol. m^{-3}, Si: 2. 2 mmol. m^{-3};
829 Figure S4b and Table S2a), characteristic of an HNLC region (Smetacek and Naqvi, 2008;
830 Quéguiner 2013). Of note, the low concentration of silicates in this station may have acted as a
831 limiting factor for the growth of diatoms. Station TARA_123 is coastal, 8 km downstream of
832 Nuku Hiva island and with a seabed depth of 1,903 m and higher chlorophyll levels ($[\text{Chl-}a]_{\text{int}}$
833 33.6 $\text{mg}\cdot\text{m}^{-2}$), indicative of a bloom. Nutrients were as elevated as in the pre-bloom HNLC area,
834 and with a particular increase of NO_{2-} around 150 m depth (1.47 $\text{mmol}\cdot\text{m}^{-3}$). Station TARA_124
835 is away from the coast, 43 km from Nuku Hiva, in even deeper water (2,414 m bottom depth),
836 and in an eddy also characterized by high chlorophyll content with respect to Station TARA_122
837 ($[\text{Chl-}a]_{\text{int}}$ 28.5 $\text{mg}\cdot\text{m}^{-2}$). The chlorophyll patch was possibly seeded near the islands and
838 transported by currents far from the coast, but sustained by the eddy dynamics and its interaction
839 with underlying water. Station TARA_125 is located 300 km downstream of the islands. The
840 chlorophyll patch was still clearly evident ($[\text{Chl-}a]_{\text{int}}$ 27.6 $\text{mg}\cdot\text{m}^{-2}$). Of note, the large NO_{2-}
841 reservoir at the base of the mixed layer (120-180 m) (Figure S4b) may indicate significant
842 biological activity, although our data are not sufficient to discriminate which is the relative
843 contribution of phyto-, zoo- and bacterio-plankton to establish the nitrite reservoirs.

The concentration of measured biologically relevant metals was generally reduced in Stations TARA_123 to TARA_125 with respect to HNLC station TARA_122 (Table S2b). The reduction of dissolved ions was particularly significant in the case of cobalt, nickel, copper and cadmium, which may be considered as a potential clue for an increased uptake of biologically available trace metals in the leeward stations, although other mechanisms cannot be ruled out. Since these metals were not limiting in the HNLC conditions, it is possible that the removal of iron limitation affected the biological pathways related to metal ion uptake in general. For more information on the oceanographic context of the Marquesas Island at the time of sampling, see Supplementary Material S2.

At the four Marquesas sampling sites the IAAs displayed dynamic patterns (Figure 7b, c; Table S1f; Supplementary Material S3). The low-iron adapted DarkRed IAA showed a progressive decrease in its prominence leeward of the islands, consistent with its negative correlations to iron at global level, while the Turquoise IAA showed increases in abundance. The Turquoise IAA is the only module containing autotrophs both positively and negatively correlated with iron, and while the latter were prominent at Station TARA_122 the former were prevalent at stations TARA_123-125 (Figure 7c). The observed changes in IAA prevalence in the Marquesas stations therefore supports a role for iron in the modulation of plankton communities in the region. Prokaryote IAAs, although not taxonomy-based, are dynamically responsive at the Marquesas Islands (Figure 2b, lower panel). Two types of response can be detected: a) the prokaryote IAAs Grey60 and Plum1 show a shift from negative to positive eigenlineage scores from TARA_122 to TARA_123, and b) the SaddleBrown, Red and SkyBlue IAAs show eigenvalue peaks in Station TARA_123.

The dynamics of the DarkRed subnetwork at the Marquesas Islands may be used to examine our previous claim that it is the module as a whole that responds to perturbation. This is a low-iron associated module, in which the autotrophic species are the most relevant and typically show negative correlations to iron at the global level, in contrast to most of the autotrophs in the other modules (Figure 7c). Indeed, only a few of the species associated to this subnetwork are present in high-iron conditions. In the context of the Marquesas Islands, most of the DarkRed-assigned OTUs were detected in the oligotrophic Station TARA_122 but not at Station TARA_123, where iron was not expected to be a limiting factor. Both richness and the relative abundance of the subnetwork decreased at Station TARA_123. We thus observe that an iron-responsive sub-

network changes its richness and abundance in a manner consistent with iron availability, rearranging the connectivity between its nodes. Additional details about the dynamics of IAAs in the Marquesas archipelago can be found in Supplementary Material S3.g.

Further analysis of the plankton communities at the Marquesas stations showed that the biomass of primary producers was around 50% higher at the leeward stations (Stations TARA_123, 124 and 125) than at HNLC Station TARA_122, with increases in diatoms, haptophytes, pelagophytes, and *Synechococcus* (Figure 8a-d; Supplementary Material S3). The higher productivity likely fueled increases in zooplankton standing stock at these three stations, in particular copepods, chaetognaths, and appendicularians (Figure 8b, e).

Eukaryotic phytoplankton diversity increased at TARA_123 (Figure 8f, Supplementary Material S3), likely favored by the intense physical dynamics (Barton et al., 2010; Biard et al., 2016). At these stations the increased number of diatoms was due principally to *Thalassiosira* and *Minutocellus* (Supplementary Material S3). Increases in haptophyte and pelagophyte abundance were due to *Phaeocystis* and *Pelagomonas*, respectively. By contrast, the community at Station TARA_122 was more characteristic of an extremely oligotrophic environment, with an abundance of Rhizaria (Biard et al., 2016), *Planktoniella* diatoms (Malviya et al., 2016), *Chrysochromulina* haptophytes (Stibor et al., 2003), and *Pelagococcus* pelagophytes (Guillou et al., 1999), as well as *Prochlorococcus* (Rusch et al., 2010).

Analysis of picocyanobacteria also revealed alterations consistent with increased iron bioavailability in the wake of the islands with respect to TARA_122 (Figure 6). For example, we observed an almost complete shift of *Synechococcus* community composition from clade CRD1 at TARA_122 to clade II at TARA_123 and TARA_124, while absolute abundances of *Prochlorococcus* HLIII and IV, previously shown to dominate in iron-depleted waters (Rusch et al., 2010; West at al., 2011), were significantly reduced (Figure 6; Supplementary Material S3 and S4).

Using transcriptomes from MMETSP together with metatranscriptomes from *Tara* Oceans (Sunagawa et al., 2015; Louca et al., 2016; Alberti et al., 2017; Carradec et al., 2018) we could further compare the qualitative shifts in genotypes highlighted above with changes in transcriptional outputs in cyanobacteria (Figure 9a, b), eukaryotic phytoplankton (Figure 9c, d, Figure S5), metazoans (Figure S6 and Supplementary Material S4), and more specifically in diatoms (Figure 9c, d, Figure S7). Importantly, *ISIP* levels were decreased in the leeward stations

(Figure 5a, Figure 9c, d), and study of gene switches proposed to be responsive to ambient iron concentrations such as ferredoxin/flavodoxin, plastocyanin/cytochrome c_6, and FBAI/FBAII (Peers and Price, 2006; Thompson et al., 2011; Allen et al., 2012; Marchetti et al., 2012; Mackey et al., 2015; Pierella Karlusich et al., 2015) revealed patterns generally consistent with increased bioavailability at Stations TARA_123-125 with respect to HNLC Station TARA_122 both in *Synechococcus* (Figure 9a, b) and in the major groups of eukaryotic phytoplankton (Figure 9c, d and Figures S5 and S7). The abundance of proteorhodopsin and ferritin genes and mRNA in diatoms were generally also consistent with this hypothesis, with decreases in proteorhodopsin transcripts and increases in ferritin in Station TARA_123 with respect to Station TARA_122 (Figure 9c, d and Figure S7d). These patterns of known iron-responsive genes provide strong support that iron bioavailability is an important driver of the phytoplankton blooms in the Marquesas Islands (Supplementary Material S4).

Furthermore, and consistently with the global analyses, *Thalassiosira* and *Pseudo-nitzschia* appear to employ different mechanisms to respond to iron in the Marquesas stations. Specifically, small ferritin-containing *Thalassiosira* cells expressing cytochrome c_6 genes increase in abundance at Station TARA_123, replacing larger Thalassiosirales genetically adapted to low iron at Station TARA_122 by their almost exclusive expression of plastocyanin with respect to cytochrome c_6 (Figure 9c, d, Figure S7; Supplementary Material S4). On the other hand, *Pseudo-nitzschia* cells with flavodoxin and plastocyanin genes are enriched in TARA_122 in comparison with TARA_123. For these two diatom genera, the investigation of the local response around the Marquesas Islands therefore corroborates their behavior within IAAs at the global level, and their compartmentalization into different groups based on *ISIP* gene abundance and expression (Figure 5) supports the hypothesis that they have evolved fundamentally different mechanisms to respond to iron resource availability.

The outcome of the taxon-specific responses summarized above and discussed more comprehensively in the Supplementary Material S4 are shifts in abundance and occurrence of taxa within IAAs that change the overall structure of the food web. Our observations also reveal novel information about the genetic strategies and specialized mechanisms employed by each taxon to cope with iron availability (Supplementary Material S4) and illustrate that these responses may ensure resilience of each IAA in a subset of conditions within a highly variable environment. Collectively, our results therefore demonstrate that the delineation of co-responsive

sub-communities at global scale can provide a valuable framework for identifying key lineages whose adaptive capacities can be compared and contrasted in specific dynamic contexts. Finally, our in-depth analysis of community structure and gene expression around the Marquesas Islands illustrates how biological data can be used to inform biogeochemical models, because neither of the models used here was able to project increased iron availability in the wake of the islands. Furthermore, while the four Marquesas stations were used in the global analysis that defined the IAAs, they did not contribute to the correlation of IAAs because of the lack of resolution of the models in this area. The module responses in the Marquesas are therefore not biased but are remarkably indicative of a change in iron bioavailability in the lee of the islands.

4. Discussion

In this study we have shown how the turnover of organisms coping with ocean variability involves a combination of ontogenetic responses driven essentially by modulation of gene expression patterns, i.e., acclimation, together with phylogenetic responses driven by changes in plankton community structure as well as different genotypes adapted to local conditions by altered copy numbers of iron responsive genes. Different organismal groups appear to use different strategies, meaning that they will not all respond over the same evolutionary timescales. The island mass effect in the wake of the Marquesas Islands leads to the selection of preferred genotypes at the community level and triggers acclimatory responses to fine-tune metabolic functioning via transcriptional responses. These local observations of the most affected organisms are consistent with IAAs identified in the global ocean, suggesting that large scale equilibria are in fact dynamic and responsive to smaller scale perturbations.

Previous studies at global scale of the effects of iron on marine plankton were focused on a specific subset of bacterial genes involved in iron metabolism using metagenomics samples from North West Atlantic, Equatorial Pacific and Indian Oceans (Toulza et al., 2012). Our current study extends this analysis because of its broader geographical coverage and the vastly expanded sequencing dataset, which has permitted us to explore both community-level and gene-level responses throughout the entire plankton community, from viruses to zooplankton. Our work thus provides an extensive global scale analysis of the different levels at which plankton

biodiversity may be impacted by iron availability, although it should not be assumed that all the responses we highlighted depend solely on iron because one single resource is very unlikely to drive the physiological and structural dynamics of a community. Nonetheless, our extensive statistical analyses suggest that the responses we define do certainly involve iron bioavailability and that the responses occur at molecular, physiological and compositional levels. Of note is the evidence of modularity in the community structure with modules of co-occurring taxa being sensitive to the resource yet displaying often contrasting strategies. This extends the results obtained by (Guidi et al., 2016) who focused on a specific process, indicating that modularity is a general feature of plankton communities, which might be related to their continuous turnover. To the extent allowed by available gene catalogs and taxonomic resolution we were able to link the sub-community responses to the molecular toolkits of the organisms but in many cases we emphasize that the response is not unequivocal but rather maps to a suite of strategies that had already been recorded previously in localized or laboratory experiments.

The complexity of the plankton ecosystem that emerges from the analysis of each IAA and their VIPs, whose dynamics have a certain degree of freedom with respect to the response of the others, indicates that there is some flexibility between the composition of primary producers and their consumers, even though the former are the organisms most directly impacted by nutrient availability. In particular, heterotrophic grazers appear to be central for responses to such bottom-up processes as nutrient acquisition. We interpret the VIP values vs. correlation to iron and community centrality as follows: that communities are assemblages of several organisms with multiple interactions among them that cannot be reduced to just a handful of opportunistic autotrophic species able to benefit from nutrient injection and that supply organic carbon to higher trophic levels. Rather, organisms respond to resource availability according to their functional traits but also modulate interactions within their communities, thus affecting their structure. These changes will nonetheless depend on the resident community, immigration from beyond, and changes in the ambient conditions. Some organisms may thrive in different contexts and therefore not be strongly dependent on iron, but rather be good exploiters of primary production stimulated by increased nutrient bioavailability; most of the VIPs are indeed consumers. Furthermore, the relatively low subnetwork centrality of these consumers may suggest that they co-occur with only a subcomponent of the other species. Finally, the nature of the modules composed of parasitic and mixotrophic organisms further suggests that recycling of

matter, e.g., through remineralization, parasitism and pathogenesis, are additional strategies within plankton communities to overcome resource limitation. Such strategies would be expected to confer further flexibility, and lead to an improved capacity to respond to sporadic bursts of favorable conditions.

Taxonomy-based network analysis for the prokaryotes did not reveal significant associations with iron bioavailability, whereas their gene subnetworks did. In accordance with a recent study based largely on *Tara* Oceans data (Louca et al., 2016), this result advocates for the use of prokaryotic functional signals rather than standard taxonomic criteria to study functional responses of prokaryotes in the global ocean, at least at the level of current taxonomic resolution. In fact, picocyanobacteria displayed a remarkable strain-dependent sensitivity to iron availability. The observations further indicate the need for a better assignation of functional taxonomy, and more studies to better characterize prokaryotic genes of relevance for interpreting the mechanistic changes in prokaryotes following perturbations in iron bioavailability. Furthermore, while standard steady-state analyses of ocean systems do not consider biological responses to perturbation per se, our approach of identifying steady-state global IAA subnetworks and then investigating their responses to local, short-term perturbation represents a promising new approach.

Comparison of the local response to an inferred iron injection in the Marquesas archipelago with the global patterns indicates that the community response to iron availability cannot be characterized by an even increase in biomass among existing components but involves a change in their relative weights reflecting their different adaptive solutions and the concurrent reorganization of the sub-communities. In other words, our results infer that the rate of supply of a resource is a factor that modulates the response of organisms and their communities.

Our analysis is based on iron distribution derived from two advanced biogeochemical models rather than from discrete measurements. This is because we considered them to be more representative than the instantaneous *in situ* measurements whose coverage is also scarce and could not be improved by our expedition, since *Tara* was not equipped to accurately perform iron concentration assessments. While this may be viewed as a limitation of our work, we provide evidence from independent data of the reliability of these estimates, thus providing a valuable demonstration of the utility of omics data as a tool to validate (and consequently improve) current models of earth system dynamics. The good correspondence between the

molecular response and the model simulations demonstrates that metatranscriptomics is now mature enough to provide an independent, biologically-based validation of ecosystem models especially when the data are scarce or hard to obtain in a reliable way. The quality and number of iron measurements are continuously improving but metatranscriptomics may anticipate and suggest the presence of biogeochemical constraints that are still undetectable with analytical methods. In addition, it could significantly integrate the formulation of processes in current ecological models because, on the long term, it can complement the missing information about organism interactions (see above) which cannot be derived from the availability of resources (e.g., Stec et al., 2017).

In conclusion, our study reinforces the results obtained in smaller-scale studies and significantly expands the suite of indicators that can be monitored to detect responses to changes in environmental conditions, from target genes to higher levels of biological organization. Our work paves the way to a suite of possible developments in experimental design and in model formulations that prompt for the improvement of statistical tools to better characterize responses at system level. Numerical simulations of ocean processes aimed at capturing the fluxes of key elements are currently based on just a handful of plankton functional types (Le Quere et al., 2005) or functional genes (Coles et al., 2017). Our results highlight the need to incorporate the response of entire plankton assemblages to more accurately determine responses at different levels, such as gene expression, gene copy numbers, or community composition. To determine the relevance of such processes omics should become a routine component of ocean observation, and we further demonstrate here that it can contribute to assessing the validity of ecosystem models by complementing biogeochemical measurements in the field and adding critical information about the actual bioavailability of nutrients, which is currently difficult to measure. Finally, the IAAs and other modules described herein provide a framework that is independent of taxonomic or functional groupings to tackle the complexity of natural communities, thus assisting our capacity to predict the responses and resilience of planktonic ecosystems to natural and human-induced perturbations.

Acknowledgements

The *Tara* Oceans consortium acknowledges the origin of samples from Stations TARA_113-125 as French Polynesia and that they were collected under Convention number 3534 (Convention relatif à la campagne de prélévements et de mésures de Tara Oceans en Polynesie Francaise) dated 16 June 2011. We thank the commitment of the following people and sponsors who made this singular expedition possible: CNRS (in particular Groupement de Recherche GDR3280**,** the Mission Pour l'Interdisciplinarité – Project MEGALODOM, and the Fédération de Recherche GO-SEE FR2022), European Molecular Biology Laboratory (EMBL), Genoscope/CEA, the French Government 'Investissements d'Avenir' programs Oceanomics (ANR-11-BTBR-0008), MEMO LIFE (ANR-10-LABX-54), PSL* Research University (ANR-11-IDEX-0001-02), and FRANCE GENOMIQUE (ANR-10-INBS-09), Fund for Scientific Research – Flanders, VIB, Stazione Zoologica Anton Dohrn, UNIMIB, ANR (projects 'PHYTBACK/ANR-2010-1709-01', POSEIDON/ANR-09-BLAN-0348, PROMETHEUS/ANR-09-PCS-GENM-217, TARA-GIRUS/ANR-09-PCS-GENM-218, SAMOSA/ANR-13-ADAP-0010, CINNAMON/ANR-17-CE02-0014-01), EU FP7 (MicroB3/No. 287589), ERC Advanced Grant Award (Diatomite: 294823), the LouisD foundation of the Institut de France, a Radcliffe Institute Fellowship from Harvard University to CB, JSPS/MEXT KAKENHI (Nos. 26430184, 16H06437, 16KT0020), The Canon Foundation (No. 203143100025), agnès b., the Veolia Environment Foundation, Region Bretagne, World Courier, Illumina, Cap L'Orient, the EDF Foundation EDF Diversiterre, FRB, the Prince Albert II de Monaco Foundation, Etienne Bourgois, the Fonds Français pour l'Environnement Mondial, the Tara schooner and its captain and crew. *Tara* Oceans would not exist without continuous support from 23 institutes (http://oceans. taraexpeditions. org). This article is contribution number XX of *Tara* Oceans.

The authors have deposited the data in the following repositories: sequencing data are archived at ENA (http://www.ebi.ac.uk/ena/) under the accession number PRJEB4352 for the metagenomics data and PRJEB6609 for the metatranscriptomics data (Carradec et al., 2018); environmental data are available at PANGAEA (https://www.pangaea.de/).

Competing interests

The authors declare that no competing interests exist.

Author Contribution

DI, FN, Fd'O and PT designed the study with input from *Tara* Oceans coordinators. DI directed the project. FN directed the field work in the Marquesas archipelago. CB wrote the paper with substantial input from MRd'A, DI, LC, and other first authors. Fd'O, PT, FN, EM, DI, HC, LG, SS, FK performed oceanographic analyses, AT and MJF provided iron concentration data from biogeochemical models, FRJV, GB and AT compared iron products from different biogeochemical models, ES, AZ, SM, JV, JL, SC, FV, AtT, CB performed analysis of eukaryotic phytoplankton, MGM, J-LJ, J-BR, SG, LC, LS, FL, TB performed analysis of metazoans and other zooplankton, JJPK, SC, HD and LG performed analysis of cyanobacteria, QC, EP, FRJV, JJPK, EV, SGA, AA, SSu, PB, PW, AV, RS, JP, GL-M, ML performed global omics analyses, FRJV, JJPK, AK, JP-Y, LT performed analysis of omics data from eukaryotic phytoplankton, AK, EP, LC, PS, SdA performed analysis of omics data from metazoans and other zooplankton, JJPK, HD and LG performed gene expression analysis of cyanobacteria, JJPK, JRB, SR, MBS, and MB performed analysis of viruses, LB, SC A-SB and DE performed WGCNA analyses, SR, FN, CD, MP, SKL, SSe and SP collected and managed *Tara* Oceans samples, LC, JJPK, MRd'A and CB assembled the manuscript. *Tara* Oceans coordinators provided a creative environment and constructive criticism throughout the study. All authors discussed the results and commented on the manuscript.

References

Alberti, A., Poulain, J., Engelen, S., Labadie, K., Romac, S., Ferrera, I., et al. (2017). Viral to metazoan marine plankton nucleotide sequences from the Tara Oceans expedition. *Scientific Data*. https://doi.org/10.1038/sdata.2017.93

Alexander, H., Rouco, M., Haley, S., Wilson, S., Karl, D., & Dyhrman, S. (2015). Functional group-specific traits drive phytoplankton dynamics in the oligotrophic ocean. *Proceedings of the National Academy of Sciences*, E5972–E5979. https://doi.org/10.1073/pnas.1518165112

Allen, A. E., Moustafa, A., Montsant, A., Eckert, A., Kroth, P. G., & Bowler, C. (2012). Evolution and functional diversification of fructose bisphosphate aldolase genes in

1117 photosynthetic marine diatoms. *Molecular Biology and Evolution*, *29*(1), 367–379.
1118 https://doi.org/10.1093/molbev/msr223

1119 Allen, A. E., Laroche, J., Maheswari, U., Lommer, M., Schauer, N., Lopez, P. J., et al. (2008).
1120 Whole-cell response of the pennate diatom Phaeodactylum tricornutum to iron starvation.
1121 *Proceedings of the National Academy of Sciences of the United States of America*, *105*(30),
1122 10438–10443. https://doi.org/10.1073/pnas.0711370105

1123 Arienzo, M., Toscano, F., Di Fraia, M., Caputi, L., Sordino, P., Guida, M., et al. (2014). An
1124 assessment of contamination of the Fusaro Lagoon (Campania Province, southern Italy) by
1125 trace metals. *Environmental Monitoring and Assessment*, *186*(9), 5731–5747.
1126 https://doi.org/10.1007/s10661-014-3816-4

1127 Aumont, O., Tagliabue, A., Bopp, L., & Gehlen, M. (2015). PISCES-v2 : an ocean
1128 biogeochemical model for carbon and. *Geoscientific Model Development*, *8*(1990), 2465–
1129 2513. https://doi.org/10.5194/gmd-8-2465-2015

1130 Barton, A. D., Dutkiewicz, S., Flierl, G., Bragg, J., & Follows, M. J. (2010). Patterns of diversity
1131 in marine phytoplankton. *Science (New York, N.Y.)*.
1132 https://doi.org/10.1126/science.1184961

1133 Bartual, S. G., Otero, J. M., Garcia-Doval, C., Llamas-Saiz, A. L., Kahn, R., Fox, G. C., & van
1134 Raaij, M. J. (2010). Structure of the bacteriophage T4 long tail fiber receptor-binding tip.
1135 *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.1011218107

1136 Battaglia, M., Olvera-Carrillo, Y., Garciarrubio, A., Campos, F., & Covarrubias, A. A. (2008).
1137 The enigmatic LEA proteins and other hydrophilins. *Plant Physiology*, *148*(1), 6–24.
1138 https://doi.org/10.1104/pp.108.120725

1139 Berline, L., Stemmann, L., Vichi, M., Lombard, F., & Gorsky, G. (2011). Impact of
1140 appendicularians on detritus and export fluxes: A model approach at DyFAMed site.
1141 *Journal of Plankton Research*, *33*(6), 855–872. https://doi.org/10.1093/plankt/fbq163

1142 Bernardes, J. S., Vieira, F. R. J., Zaverucha, G., & Carbone, A. (2015). A multi-objective
1143 optimization approach accurately resolves protein domain architectures. *Bioinformatics*,
1144 *32*(3), 345–353. https://doi.org/10.1093/bioinformatics/btv582

1145 Bernardes, J., Zaverucha, G., Vaquero, C., & Carbone, A. (2016). Improvement in Protein
1146 Domain Identification Is Reached by Breaking Consensus, with the Agreement of Many
1147 Profiles and Domain Co-occurrence. *PLoS Computational Biology*, *12*(7), 1–39.
1148 https://doi.org/10.1371/journal.pcbi.1005038

1149 Biard, T., Stemmann, L., Picheral, M., Mayot, N., Vandromme, P., Hauss, H., et al. (2016). In
1150 situ imaging reveals the biomass of giant protists in the global ocean. *Nature*, *532*(7600),
1151 504–507. https://doi.org/10.1038/nature17652

1152 Blain, S., Bonnet, S., & Guieu, C. (2008). Dissolved iron distribution in the tropical and sub
1153 tropical South Eastern Pacific. *Biogeosciences*, *5*(1), 269–280. https://doi.org/10.5194/bgd-
1154 4-2845-2007

1155 Bonnain, C., Breitbart, M., & Buck, K. N. (2016). The Ferrojan Horse Hypothesis: iron-virus
1156 interactions in the ocean. *Frontiers in Marine Science*, *3*(June), 82.
1157 https://doi.org/10.3389/fmars.2016.00082

1158 Bork, P., Bowler, C., Vargas, C. de, Gorsky, G., Karsenti, E., & Wincker, P. (2015). Tara Oceans
1159 studies plankton at planetary scale. *Science*, *348*(6237), 873.
1160 https://doi.org/10.1126/science.aac5605

1161 Botebol, H., Lesuisse, E., Šuták, R., Six, C., Lozano, J.-C., Schatt, P., et al. (2015). Central role
1162 for ferritin in the day/night regulation of iron homeostasis in marine phytoplankton.
1163 *Proceedings of the National Academy of Sciences*, *112*(47), 14652–14657.
1164 https://doi.org/10.1073/pnas.1506074112

1165 Boyd, P. W., Jickells, T., Law, C. S., Blain, S., Boyle, E. a, Buesseler, K. O., et al. (2007).
1166 Mesoscale iron enrichment experiments 1993-2005: synthesis and future directions. *Science*
1167 *(New York, N.Y.)*, *315*(5812), 612–617. https://doi.org/10.1126/science.1131669

1168 Browning, C., Shneider, M. M., Bowman, V. D., Schwarzer, D., & Leiman, P. G. (2012). Phage
1169 pierces the host cell membrane with the iron-loaded spike. *Structure*.
1170 https://doi.org/10.1016/j.str.2011.12.009

1171 Brum, J. R., Ignacio-espinoza, J. C., Roux, S., Doulcier, G., Acinas, S. G., Alberti, A., &
1172 Chaffron, S. (2015). Patterns and ecological drivers of ocean viral communities. *Science*,
1173 *348*(6237), 1261498-1–11. https://doi.org/10.1126/science.1261498

1174 Bundy, J. G., & Kille, P. (2014). Metabolites and metals in Metazoa--what role do
1175 phytochelatins play in animals? *Metallomics : Integrated Biometal Science*, *6*(9), 1576–82.
1176 https://doi.org/10.1039/c4mt00078a

1177 Campbell, L., Nolla, H. A., & Vaulot, D. (1994). The importance of Prochlorococcus to
1178 community structure in the central North Pacific Ocean. *Limnol. Oceanogr.*, *39*(4), 954–
1179 961.

1180 Carradec, Q., Pelletier, E., Da Silva, C., Alberti, A., Seeleuthner, Y., Blanc-Mathieu, R., et al.
1181 (2018). A global ocean atlas of eukaryotic genes. *Nature Communications*.
1182 https://doi.org/10.1038/s41467-017-02342-1

1183 Chappell, P. D., Whitney, L. P., Wallace, J. R., Darer, A. I., Jean-Charles, S., & Jenkins, B. D.
1184 (2015). Genetic indicators of iron limitation in wild populations of Thalassiosira oceanica
1185 from the northeast Pacific Ocean. *Isme J*, *9*(3), 592–602.
1186 https://doi.org/10.1038/ismej.2014.171

1187 Chen, X., Wakeham, S. G., & Fisher, N. S. (2011). Influence of iron on fatty acid and sterol
1188 composition of marine phytoplankton and copepod consumers. *Limnology and*
1189 *Oceanography*, *56*(2), 716–724. https://doi.org/10.4319/lo.2011.56.2.0716

1190 Chow, C.-E. T., & Suttle, C. A. (2015). Biogeography of Viruses in the Sea. *Annual Review of*
1191 *Virology*, *2*(1), 41–66. https://doi.org/10.1146/annurev-virology-031413-085540

1192 Claustre, H., Sciandra, A., & Vaulot, D. (2008). Introduction to the special section bio-optical
1193 and biogeochemical conditions in the South East Pacific in late 2004: The BIOSOPE
1194 program. *Biogeosciences*. https://doi.org/10.5194/bg-5-679-2008

1195 Coale, K. H., Fitzwater, S. E., Gordon, R. M., Johnson, K. S., & Barber, R. T. (1996). Control of
1196 community growth and export production by upwelled iron in the equatorial Pacific Ocean
1197 Kenneth. *Nature*. https://doi.org/10.1038/379621a0

1198 Cohen, N. R., Brzezinski, M. A., Twining, K. T. B. S., Ellis, K. A., Lampe, R. H., McNair, H., et
1199 al. (2017). Diatom transcriptional and physiological responses to changes in iron
1200 bioavailability across ocean provinces. *Frontiers in Marine Science* , *4*(November), 1–20.
1201 https://doi.org/10.3389/fmars.2017.00360

1202 Cohen, N. R., Mann, E., Stemple, B., Moreno, C. M., Rauschenberg, S., Jacquot, J., E., Sunda,
1203 W. G., Twining, B., S., Marchetti, A. (2018). Iron storage capacities and associated ferritin
1204 gene expression among marine diatoms. *Limnol. Oceanogr*, *63*(4).
1205 https://doi.org/10.1002/lno.10800

1206 Coles, V. J., Stukel, M. R., Brooks, M. T., Burd, A., Crump, B. C., Moran, M. A., et al. (2017).
1207 Ocean biogeochemistry modeled with emergent trait-based genomics. *Science*,
1208 *1154*(December), 1–26. https://doi.org/10.1126/science.aan5712

1209 Crans, D. C., Smee, J. J., Gaidamauskas, E., & Yang, L. (2004). The Chemistry and
1210 Biochemistry of Vanadium and the Biological Activities Exerted by Vanadium Compounds.
1211 *Chemical Reviews*, *104*(2), 849–902. https://doi.org/10.1021/cr020607t

1212 Dandonneau, Y., & Charpy, L. (1985). An empirical approach to the island mass effect in the
1213 south tropical Pacific based on sea surface chlorophyll concentrations. *Deep Sea Research*
1214 *Part A, Oceanographic Research Papers*, *32*(6). https://doi.org/10.1016/0198-
1215 0149(85)90074-3

1216 de Vargas, C., Audic, S., Henry, N., Decelle, J., & Mahé, F. (2015). Eukaryotic plankton
1217 diversity in the sunlit ocean. *Science*, *348*(MAY), 1–12. https://doi.org/10.1007/s13398-
1218 014-0173-7.2

1219 De Vos, C. H., Vonk, M. J., Vooijs, R., & Schat, H. (1992). Glutathione Depletion Due to
1220 Copper-Induced Phytochelatin Synthesis Causes Oxidative Stress in Silene cucubalus. *Plant*
1221 *Physiology*, *98*(3), 853–858. https://doi.org/10.1104/pp.98.3.853

1222 Dolan, J. R., Ritchie, M. E., & Ras, J. (2007). The “neutral” community structure of planktonic
1223 herbivores, tintinnid ciliates of the microzooplankton, across the SE Tropical Pacific Ocean.
1224 *Biogeosciences Discussions*, *4*, 561–593. https://doi.org/10.5194/bgd-4-561-2007

Dupont, C. L., Mccrow, J. P., Valas, R., Moustafa, A., Walworth, N., Goodenough, U., et al. (2015). Genomes and gene expression across light and productivity gradients in eastern subtropical Pacific microbial communities. *ISME Journal*, *9*(5), 1076–1092. https://doi.org/10.1038/ismej.2014.198

Durkin, C.A., Marchetti, A., Bender, S.J., Truong, T., Morales, R., Mock, T., Armbrust, V. E., (2012). Frustule-related gene transcription and the influence of diatom community composition on silica precipitation in an iron-limited environment. *Limnol. Oceanogr. 57,* 1619–1633. doi:10.4319/lo.2012.57.6.1619

Durkin, C. A., Koester, J. A., Bender, S. J., Armbrust, E. V., & Kroth, P. (2016). The evolution of silicon transporters in diatoms. *Journal of Phycology*, *52*(5), 716–731. https://doi.org/10.1111/jpy.12441

Farrant, G. K., Doré, H., Cornejo-Castillo, F. M., Partensky, F., Ratin, M., Ostrowski, M., et al. (2016). Delineating ecologically significant taxonomic units from global patterns of marine picocyanobacteria. *Proceedings of the National Academy of Sciences*, *113*(24), E3365–E3374. https://doi.org/10.1073/pnas.1524865113

Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., et al. (2016). The Pfam protein families database: Towards a more sustainable future. *Nucleic Acids Research*, *44*(D1), D279–D285. https://doi.org/10.1093/nar/gkv1344

Follows, M. J., Dutkiewicz, S., Grant, S., & Chisholm, S. W. (2007). Emergent Biogeography of Microbial. *Science*, (C), 1843–1847. https://doi.org/10.1126/science.1138544

Freeland, H. J., & Cummins, P. F. (2005). Argo: A new tool for environmental monitoring and assessment of the world's oceans, an example from the N.E. Pacific. *Progress in Oceanography*, *64*(1), 31–44. https://doi.org/10.1016/j.pocean.2004.11.002

Gómez, F., Claustre, H., Raimbault, P., & Souissi, S. (2007). Two High-Nutrient Low-Chlorophyll phytoplankton assemblages: the tropical central Pacific and the offshore Perú-Chile Current. *Biogeosciences Discussions*, *4*(3), 1535–1554. https://doi.org/10.5194/bgd-4-1535-2007

1252 Gong, W., Browne, J., Hall, N., Schruth, D., Paerl, H., & Marchetti, A. (2016). Molecular
1253 insights into a dinoflagellate bloom. *The ISME Journal*, 1–14.
1254 https://doi.org/10.1038/ismej.2016.129

1255 Gorsky, G., Chrétiennot-Dinet, M. J., Blanchot, J., & Palazzoli, I. (1999). Picoplankton and
1256 nanoplankton aggregation by appendicularians: Fecal pellet contents of Megalocercus
1257 huxleyi in the equatorial Pacific. *Journal of Geophysical Research*, *104*(C2), 3381–3390.
1258 https://doi.org/10.1029/98jc01850

1259 Gove, J. M., McManus, M. A., Neuheimer, A. B., Polovina, J. J., Drazen, J. C., Smith, C. R., et
1260 al. (2015). Ocean Oases: near-island biological hotspots in barren ocean basins. *Nature*
1261 *Communications*, *7*, Article number: 10581. https://doi.org/10.1038/ncomms10581

1262 Graff van Creveld, S., Rosenwasser, S., Levin, Y., & Vardi, A. (2016). Chronic iron limitation
1263 confers transient resistance to oxidative stress in marine diatoms. *Plant Physiology*.
1264 https://doi.org/10.1104/pp.16.00840

1265 Greene, R.M., Geider, R.J., Falkowksi, P.G. (1991). Effect of iron limitation on photosynthesis
1266 in a marine diatom. *Limnology and Oceanography, 36(8)*
1267 **https://doi.org/10.4319/lo.1991.36.8.1772**

1268 Grob, C., Ulloa, O., Claustre, H., Huot, Y., Alarcón, G., & Marie, D. (2007). Contribution of
1269 picoplankton to the total particulate organic carbon concentration in the eastern South
1270 Pacific. *Biogeosciences*, *4*(5), 837–852. https://doi.org/10.5194/bg-4-837-2007

1271 Groussman, R. D., Parker, M. S., & Armbrust, E. V. (2015). Diversity and evolutionary history
1272 of iron metabolism genes in diatoms. *PLoS ONE*, *10*(6), 1–25.
1273 https://doi.org/10.1371/journal.pone.0129081

1274 Guidi, L., Gorsky, G., Claustre, H., Miquel, J. C., Picheral, M., & Stemmann, L. (2008).
1275 Distribution and fluxes of aggregates >100 μm in the upper kilometer of the South-
1276 Eastern Pacific. *Biogeosciences*, *5*(5), 1361–1372. https://doi.org/10.5194/bg-5-1361-2008

1277 Guidi, L., Chaffron, S., Bittner, L., Eveillard, D., Larhlimi, A., Roux, S., et al. (2016). Plankton
1278 networks driving carbon export in the oligotrophic ocean. *Nature*, *532*, 465–470.
1279 https://doi.org/10.1038/nature16942

1280 Guillou, L., Moon-Van Der Staay, S. Y., Claustre, H., Partensky, F., & Vaulot, D. (1999).
1281 Diversity and abundance of Bolidophiceae (Heterokonta) in two oceanic regions. *Appl.*
1282 *Environ. Microbiol.*, *65*(10), 4528–4536.

1283 Guindon S., Dufayard J.F., Lefort V., Anisimova M., Hordijk W., & Gascuel O. (2010). New
1284 Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the
1285 Performance of PhyML 3.0. *Systematic Biology*, 59(3), 307-21, 2010.
1286 https://doi.org/10.1093/sysbio/syq010

1287 Karsenti, E., Acinas, S. G., Bork, P., Bowler, C., de Vargas, C., Raes, J., et al. (2011). A holistic
1288 approach to marine Eco-systems biology. *PLoS Biology*, *9*(10), 7–11.
1289 https://doi.org/10.1371/journal.pbio.1001177

1290 Katoh, K., & Daron, M.S. (2013) MAFFT multiple sequence alignment software version 7:
1291 improvements in performance and usability. *Molecular Biology and Evolution* 30(4), 772-
1292 780. https://doi.org/10.1093/molbev/mst010

1293 Kazamia, E., Sutak, R., Paz-Yepes, J., Dorrell, R. G., Vieira, F. R. J., Mach, J., et al. (2018).
1294 Endocytosis-mediated siderophore uptake as a strategy for Fe acquisition in diatoms.
1295 *Science Advances*, *4*(5), eaar4536. https://doi.org/10.1126/sciadv.aar4536

1296 Khoshnood, B., Dacklin, I., & Grabbe, C. (2016). Urm1: An essential regulator of JNK signaling
1297 and oxidative stress in Drosophila melanogaster. *Cellular and Molecular Life Sciences*,
1298 *73*(9), 1939–1954. https://doi.org/10.1007/s00018-015-2121-x

1299 Lampe, R.H Cohen, N. R., Ellis, K. A., Bruland, K.W., Maldonado M.T. , Peterson, T.D., Till,
1300 C.P., Brzezinski. M.A., Bargu, S., Thamatrakoln, K., Kuzminov, F.I., Twining, B.S.,
1301 Marchetti, A. (2018) Divergent gene expression among phytoplankton taxa in response to
1302 upwelling. Environmental Microbiology, 20(8) https://doi.org/10.1111/1462-2920.14361

1303 Lane, T. W., & Morel, F. M. M. (2000). A biological function for cadmium in marine diatoms.
1304 *Proceedings of the National Academy of Sciences*, *97*(9), 4627–4631.
1305 https://doi.org/10.1073/pnas.090091397

Langfelder, P., & Horvath, S. (2007). Eigengene networks for studying the relationships between co-expression modules. *BMC Systems Biology*, *1*(1), 54. https://doi.org/10.1186/1752-0509-1-54

Le Quere, C., Harrison, S. P., Colin Prentice, I., Buitenhuis, E. T., Aumont, O., Bopp, L., et al. (2005). Ecosystem dynamics based on plankton functional types for global ocean biogeochemistry models\rdoi:10.1111/j.1365-2486.2005.1004.x. *Global Change Biology*, *11*(11), 2016–2040. https://doi.org/10.1111/j.1365-2486.2005.01004.x

Legeckis, R., Brown, C. W., Bonjean, F., & Johnson, E. S. (2004). The influence of tropical instability waves on phytoplankton blooms in the wake of the Marquesas Islands during 1998 and on the currents observed during the drift of the Kon-Tiki in 1947. *Geophysical Research Letters*, *31*(23), 1–4. https://doi.org/10.1029/2004GL021637

Lommer M., Specht M., Roy A.S., Kraemer L., Andreson R., Gutowska M.A., Wolf J., Bergner S.V., Schilhabel M.B., Klostermeier U.C., Beiko R.G., Rosenstiel P., Hippler M., LaRoche J. (2012). Genome and low-iron response of an oceanic diatom adapted to chronic iron limitation. *Genome Biology, 13(7):*R66 doi: 10.1186/gb-2012-13-7-r66.

Lommer, M., Roy, A.-S., Schilhabel, M., Schreiber, S., Rosenstiel, P., & LaRoche, J. (2010). Recent transfer of an iron-regulated gene from the plastid to the nuclear genome in an oceanic diatom adapted to chronic iron limitation. *BMC Genomics*, *11*(1), 718. https://doi.org/10.1186/1471-2164-11-718

Louca, S., Parfrey, L. W., & Doebeli, M. (2016). Decoupling function and taxonomy in the global ocean microbiome. *Science*, *353*(6305), 1272–1277.

Mackey, K. R. M., Post, A. F., McIlvin, M. R., Cutter, G. A., John, S. G., & Saito, M. A. (2015). Divergent responses of Atlantic coastal and oceanic Synechococcus to iron limitation. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(32), 9944–9. https://doi.org/10.1073/pnas.1509448112

Mahowald, N. M., Baker, A. R., Bergametti, G., Brooks, N., Duce, R. A., Jickells, T. D., et al. (2005). Atmospheric global dust cycle and iron inputs to the ocean. *Global Biogeochemical Cycles*, *19*(4). https://doi.org/10.1029/2004GB002402

1335 Malviya, S., Scalco, E., Audic, S., Vincent, F., Veluchamy, A., Poulain, J., et al. (2016). Insights
1336 into global diatom distribution and diversity in the world's ocean. *Proceedings of the*
1337 *National Academy of Sciences*, *113*(11), E1516–E1525.
1338 https://doi.org/10.1073/pnas.1509523113

1339 Marchetti, a., Schruth, D. M., Durkin, C. a., Parker, M. S., Kodner, R. B., Berthiaume, C. T., et
1340 al. (2012). Comparative metatranscriptomics identifies molecular bases for the
1341 physiological responses of phytoplankton to varying iron availability. *Proceedings of the*
1342 *National Academy of Sciences*, *109*(6), E317–E325.
1343 https://doi.org/10.1073/pnas.1118408109

1344 Marchetti, A., Catlett, D., Hopkinson, B. M., Ellis, K., & Cassar, N. (2015). Marine diatom
1345 proteorhodopsins and their potential role in coping with low iron availability. *ISME*
1346 *Journal*, *9*(12), 2745–2748. https://doi.org/10.1038/ismej.2015.74

1347 Marchetti, A., Juneau, P., Whitney, F. A., Wong, C. S., & Harrison, P. J. (2006). Phytoplankton
1348 processes during a mesoscale iron enrichment in the NE subarctic Pacific: Part II-Nutrient
1349 utilization. *Deep-Sea Research Part II: Topical Studies in Oceanography*, *53*(20–22),
1350 2114–2130. https://doi.org/10.1016/j.dsr2.2006.05.031

1351 Marchetti, A., Moreno, C. M., Cohen, N. R., Oleinikov, I., Twining, B. S., Armbrust, E. V., &
1352 Lampe, R. H. (2017). Development of a molecular-based index for assessing iron status in
1353 bloom-forming pennate diatoms. *Journal of Phycology*, 1–13.
1354 https://doi.org/10.1111/jpy.12539

1355 Marchetti, A., Parker, M. S., Moccia, L. P., Lin, E. O., Arrieta, A. L., Ribalet, F., et al. (2009).
1356 Ferritin is used for iron storage in bloom-forming marine pennate diatoms. *Nature*,
1357 *457*(7228), 467–470. https://doi.org/10.1038/nature07539

1358 Martin, P., Van Der Loeff, M. R., Cassar, N., Vandromme, P., D'Ovidio, F., Stemmann, L., et al.
1359 (2013). Iron fertilization enhanced net community production but not downward particle
1360 flux during the Southern Ocean iron fertilization experiment LOHAFEX. *Global*
1361 *Biogeochemical Cycles*, *27*(3), 871–881. https://doi.org/10.1002/gbc.20077

1362 Martinez, E., & Maamaatuaiahutapu, K. (2004). Island mass effect in the Marquesas Islands:
1363 Time variation. *Geophysical Research Letters*, *31*(18), 1–4.
1364 https://doi.org/10.1029/2004GL020682

1365 Masquelier, S., & Vaulot, D. (2007). Distribution of micro-organisms along a transect in the
1366 South-East Pacific Ocean (BIOSOPE cruise) from epifluorescence microscopy.
1367 *Biogeosciences Discussions*, *4*(4), 2667–2697. https://doi.org/10.5194/bgd-4-2667-2007

1368 Mawji, E., Schlitzer, R., Dodas, E. M., Abadie, C., Abouchami, W., Anderson, R. F., et al.
1369 (2014). The GEOTRACES Intermediate Data Product 2014. *Marine Chemistry*, *177*, 1–8.
1370 https://doi.org/10.1016/j.marchem.2015.04.005

1371 McQuaid, J. B., Kustka, A. B., Oborník, M., Horák, A., McCrow, J. P., Karas, B. J., et al. (2018).
1372 Carbonate-sensitive phytotransferrin controls high-affinity iron uptake in diatoms. *Nature*.
1373 https://doi.org/10.1038/nature25982

1374 Menemenlis, B. D., Campin, J., Heimbach, P., Hill, C., & Lee, T. (2008). ECCO2 : High
1375 Resolution Global Ocean and Sea Ice Data Synthesis. *Mercator Ocean Quarterly*
1376 *Newsletter*, *31*(October), 13–21.

1377 Mevik, B.-H., & Wehrens, R. (2007). The pls Package: Principle Component and Partial Least
1378 Squares Regression in R. *Journal of Statistical Software*, *18*(2), 1–24.
1379 https://doi.org/10.1002/wics.10

1380 Mock, T., Samanta, M. P., Iverson, V., Berthiaume, C., Robison, M., Holtermann, K., et al.
1381 (2008). Whole-genome expression profiling of the marine diatom Thalassiosira pseudonana
1382 identifies genes involved in silicon bioprocesses. *Proceedings of the National Academy of*
1383 *Sciences*, *105*(5), 1579–1584. https://doi.org/10.1073/pnas.0707946105

1384 Morrissey, J., Sutak, R., Paz-Yepes, J., Tanaka, A., Moustafa, A., Veluchamy, A., et al. (2015).
1385 A novel protein, ubiquitous in marine phytoplankton, concentrates iron at the cell surface
1386 and facilitates uptake. *Current Biology*, *25*(3), 364–371.
1387 https://doi.org/10.1016/j.cub.2014.12.004

1388 Peers, G., & Price, N. M. N. (2006). Copper-containing plastocyanin used for electron transport
1389 by an oceanic diatom. *Nature*, *441*(7091), 341–4. https://doi.org/10.1038/nature04630

Pesant, S., Not, F., Picheral, M., Kandels-Lewis, S., Le Bescot, N., Gorsky, G., et al. (2015). Open science resources for the discovery and analysis of Tara Oceans data. *Scientific Data*, *2*(Lmd), 150023. https://doi.org/10.1038/sdata.2015.23

Pfaffen, S., Bradley, J. M., Abdulqadir, R., Firme, M. R., Moore, G. R., Brun, N. E. L., & Murphy, M. E. P. (2015). A diatom ferritin optimized for iron oxidation but not iron storage. *Journal of Biological Chemistry*, *290*(47), 28416–28427. https://doi.org/10.1074/jbc.M115.669713

Picheral, M., Guidi, L., Stemmann, L., Karl, D. M., Iddaoud, G., & Gorsky, G. (2010). The Underwater Vision Profiler 5: An advanced instrument for high spatial resolution studies of particle size spectra and zooplankton. *Limnology and Oceanography: Methods*, *8*(9), 462–473. https://doi.org/10.4319/lom.2010.8.462

Pierella Karlusich, J. J., Ceccoli, R. D., Graña, M., Romero, H., & Carrillo, N. (2015). Environmental selection pressures related to iron utilization are involved in the loss of the flavodoxin gene from the plant genome. *Genome Biology and Evolution*, *7*(3), 750–767. https://doi.org/10.1093/gbe/evv031

Probert, I., Siano, R., Poirier, C., Decelle, J., Biard, T., Tuji, A., et al. (2014). Brandtodinium gen. nov. and B.nutricula comb. Nov. (Dinophyceae), a dinoflagellate commonly found in symbiosis with polycystine radiolarians. *Journal of Phycology*, *50*(2), 388–399. https://doi.org/10.1111/jpy.12174

Quéguiner, B. (2013). Iron fertilization and the structure of planktonic communities in high nutrient regions of the Southern Ocean. *Deep-Sea Research Part II: Topical Studies in Oceanography*, *90*, 43–54. https://doi.org/10.1016/j.dsr2.2012.07.024

Ras, J., Claustre, H., & Uitz, J. (2008). Spatial variability of phytoplankton pigment distributions in the Subtropical South Pacific Ocean: comparison between in situ and predicted data. *Biogeosciences*, *5*, 353–369. https://doi.org/10.5194/bgd-4-3409-2007

Roux, S., Brum, J. R., Dutilh, B. E., Sunagawa, S., Duhaime, M. B., Loy, A., et al. (2016). Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. *Nature*, *537*(7622), 689–693. https://doi.org/10.1038/nature19366

Rusch, D. B., Martiny, A. C., Dupont, C. L., Halpern, A. L., & Venter, J. C. (2010). Characterization of Prochlorococcus clades from iron-depleted oceanic regions. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(37), 16184–16189. https://doi.org/10.1073/pnas.1009513107

Signorini, S. R., McClain, C. R., & Dandonneau, Y. (1999). Mixing and phytoplankton bloom in the wake of the Marquesas Islands. *Geophysical Research Letters*, *26*(20), 3121–3124. https://doi.org/10.1029/1999GL010470

Smetacek, V., & Naqvi, S. W. a. (2008). The next generation of iron fertilization experiments in the Southern Ocean. *Philosophical Transactions of the Royal Society A*, *366*(August), 3947–3967. https://doi.org/10.1098/rsta.2008.0144

Sohm, J. A., Ahlgren, N. A., Thomson, Z. J., Williams, C., Moffett, J. W., Saito, M. A., et al. (2016). Co-occurring Synechococcus ecotypes occupy four major oceanic regimes defined by temperature, macronutrients and iron. *ISME Journal*, *10*(2), 333–345. https://doi.org/10.1038/ismej.2015.115

Stec, K. F., Caputi, L., Buttigieg, P. L., D'Alelio, D., Ibarbalz, F. M., Sullivan, M. B., et al. (2017). Modelling plankton ecosystems in the meta-omics era. Are we ready? *Marine Genomics*, *32*, 1–17. https://doi.org/10.1016/j.margen.2017.02.006

Stemmann, L., Eloire, D., Sciandra, A., Jackson, G., Guidi, L., Picheral, M., & Gorsky, G. (2008). Volume distribution for particles between 3.5 to 2000 μm in the upper 200 m region of the South Pacific Gyre. *Biogeosciences Discussions*, *5*, 299–310. https://doi.org/10.5194/bgd-4-3377-2007

Stibor, H., & Sommer, U. (2003). Mixotrophy of a photosynthetic flagellate viewed from an optimal foraging perspective. *Protist*, *154*(1), 91–98. https://doi.org/10.1078/143446103764928512

Sunagawa, S., Coelho, L. P., Chaffron, S., Kultima, J. R., Labadie, K., Salazar, G., et al. (2015). Structure and function of the global ocean microbiome. *Science*, *348*(6237), 1261359. https://doi.org/10.1126/science.1261359

1445 Tagliabue, A., Aumont, O., DeAth, R., Dunne, J. P., Dutkiewicz, S., Galbraith, E., et al. (2016).
1446 How well do global ocean biogeochemistry models simulate dissolved iron distributions?
1447 *Global Biogeochemical Cycles*, *30*, 149–174.
1448 https://doi.org/10.1002/2015GB005289.Received

1449 Tagliabue, A., Mtshali, T., Aumont, O., Bowie, A. R., Klunder, M. B., Roychoudhury, A. N., &
1450 Swart, S. (2012). A global compilation of dissolved iron measurements: Focus on
1451 distributions and processes in the Southern Ocean. *Biogeosciences*, *9*(6), 2333–2349.
1452 https://doi.org/10.5194/bg-9-2333-2012

1453 Tagliabue, A., Bowie, A. R., Philip, W., Buck, K. N., Johnson, K. S., & Saito, M. A. (2017). The
1454 integral role of iron in ocean biogeochemistry. *Nature*, *543*(7643), 51–59.
1455 https://doi.org/10.1038/nature21058

1456 Tam, W., Pell, L. G., Bona, D., Tsai, A., Dai, X. X., Edwards, A. M., et al. (2013). Tail tip
1457 proteins related to bacteriophage λ gpL coordinate an iron-sulfur cluster. *Journal of*
1458 *Molecular Biology*. https://doi.org/10.1016/j.jmb.2013.03.032

1459 Thompson, A. W., Huang, K., Saito, M. A., & Chisholm, S. W. (2011). Transcriptome response
1460 of high- and low-light-adapted Prochlorococcus strains to changing iron availability. *The*
1461 *ISME Journal*, *5*(10), 1580–94. https://doi.org/10.1038/ismej.2011.49

1462 Toulza, E., Tagliabue, A., Blain, S., & Piganeau, G. (2012). Analysis of the global ocean
1463 sampling (GOS) project for trends in iron uptake by surface ocean microbes. *PLoS ONE*,
1464 *7*(2). https://doi.org/10.1371/journal.pone.0030931

1465 Uitz, J., Claustre, H., Morel, A., & Hooker, S. B. (2006). Vertical distribution of phytoplankton
1466 communities in open ocean: An assessment based on surface chlorophyll. *Journal of*
1467 *Geophysical Research: Oceans*, *111*(8). https://doi.org/10.1029/2005JC003207

1468 Vandeputte, D., Kathagen, G., D'Hoe, K., Vieira-Silva, S., Valles-Colomer, M., Sabino, J., et al.
1469 (2017). Quantitative microbiome profiling links gut community variation to microbial load.
1470 *Nature*. https://doi.org/10.1038/nature24460

Villar, E., Farrant, G. K., Follows, M., Garczarek, L., Speich, S., Audic, S., et al. (2015). Environmental characteristics of Agulhas rings affect interocean plankton transport. *Science*, *348*(6237), 1–12. https://doi.org/10.1126/science.1261447

West, N. J., Lebaron, P., Strutton, P. G., & Suzuki, M. T. (2011). A novel clade of Prochlorococcus found in high nutrient low chlorophyll waters in the South and Equatorial Pacific Ocean. *The ISME Journal*, *5*(6), 933–944. https://doi.org/10.1038/ismej.2010.186

Whitney, L. A. P., Lins, J. J., Hughes, M. P., Wells, M. L., Dreux Chappell, P., & Jenkins, B. D. (2011). Characterization of putative iron responsive genes as species-specific indicators of iron stress in thalassiosiroid diatoms. *Frontiers in Microbiology*, *2*(NOV), 1–14. https://doi.org/10.3389/fmicb.2011.00234

Xing, X., Morel, A., Claustre, H., D'Ortenzio, F., & Poteau, A. (2012). Combined processing and mutual interpretation of radiometry and fluorometry from autonomous profiling Bio-Argo floats: 2. Colored dissolved organic matter absorption retrieval. *Journal of Geophysical Research: Oceans*, *117*(4), 1–14. https://doi.org/10.1029/2011JC007632

Yoshinaga, R., Niwa-Kubota, M., Matsui, H., & Matsuda, Y. (2014). Characterization of iron-responsive promoters in the marine diatom Phaeodactylum tricornutum. *Marine Genomics*, *16*(1), 55–62. https://doi.org/10.1016/j.margen.2014.01.005

Yuasa, T., Horiguchi, T., Mayama, S., & Takahashi, O. (2016). Gymnoxanthella radiolariae gen. et sp. nov. (Dinophyceae), a dinoflagellate symbiont from solitary polycystine radiolarians. *Journal of Phycology*, *52*(1), 89–104. https://doi.org/10.1111/jpy.12371

Figures

Figure 1: Comparison of ECCO2-DARWIN, PISCES2 iron estimates with observed data and expression of diatom *ISIP* genes at *Tara* Oceans stations. Maps of (a) annual average iron concentrations from the ECCO2-DARWIN model (57 stations at surface), (b) from the PISCES2 model (83 stations at surface, 44 of which also at deep chlorophyll maximum depth), and (c) from the observed data where it was available at less than 2 degrees radius distance from locations of the *Tara* Oceans sampling sites (20 stations at surface, 16 of which also at deep chlorophyll maximum depth). Each circle corresponds to a sampling site, where the upper semicircle is filled according to the surface iron concentration while the lower semicircle is filled according to the deep chlorophyll maximum depth where available. Color scale indicates dissolved iron concentrations expressed in nM. (d) Biogeographical pattern of diatom *ISIP* gene expression. The circle colors represent iron concentration estimates at each *Tara* Oceans sampling site according to PISCES2 model (Table S1c). The abundance of *ISIP* transcripts was normalized by the total abundance of all diatom unigenes at each station, and the corresponding values are represented by the circle area. Boxes indicate the Marquesas Islands sampling area.

Figure 2: **Planktonic Iron-Associated Assemblages (IAAs) in the global ocean and in the Marquesas Islands stations.** (a) Description of eukaryotic modules associated with iron. Relative abundances and co-occurrences of eukaryotic lineages were used to decipher modules. Four modules can predict iron with high accuracy: Black, DarkRed, Turquoise, and Yellow. For each IAA, lineages are associated to their score of centrality (x axis), to their correlation with iron concentrations (y axis), and their VIP score (circle area). Representative lineages within each module are emphasized by circles and named (C: Copepoda, B: Bacillariophyta, R: Rhizaria). (b) Upper panel: contribution of *Tara* Oceans stations to the global variance of IAAs of eukaryotic lineages. For each IAA, we represent the projection of stations on the first principal component (upper panel). Lower panel: projection of the relative contribution of the *Tara* Oceans stations to the global variance of iron-associated prokaryotic gene assemblages, as revealed by WGCNA. For each prokaryotic gene module associated with iron (from top to bottom: Grey60, Plum1, Red, SkyBlue, and SaddleBrown), we represent the projection of stations on the first principal component, proportional to triangle sizes for each module. The behavior of each IAA in the Marquesas archipelago stations is shown in the inset.

Figure 3: **Correlation analysis between absolute cell abundance of marine picocyanobacterial clades and iron concentration estimates from PISCES2 model in surface waters**. Only statistically significant correlations are displayed (p-value<0. 05). Spearman correlation coefficients and p-values are indicated. The cell abundance for each cyanobacterial clade was assessed combining *petB* marker gene counts with flow cytometry determinations using the method published by (Vandeputte et al., 2017).

Figure 4: *Tara* Oceans metagenome survey in surface waters for oceanic phages containing putative iron-containing structural proteins. (a) Representation of protein domain architecture of viral tail proteins with putative iron-binding HxH motifs, the HMM logos for the HxH motifs identified in the corresponding *Tara* Oceans viral unigenes, and the biogeographical distribution of the corresponding viral contigs. In the map, the circle colors represent iron concentration estimates at each sampling site according to PISCES2 biogeochemical model (Table S1a), and the circle areas represent the cumulated normalized coverage of the viral contigs of interest. (b) Equivalent analysis for viral spike proteins with putative iron-binding HxH motifs. (c) Equivalent analysis for viral tail tip proteins with $CX_{8-13}CX_{22-23}CX_6C$ motif involved in 4Fe-4S cluster binding.

Figure 5: Abundance and expression of diatom *ISIP* genes with respect to iron concentration estimates. (a) 2D scatter plots correspond to the correlation between gene abundance and iron (left), gene expression and iron (middle), and abundance and expression of *ISIP* genes (right). Pearson correlation coefficients (pcc) and p-values are indicated in blue. Iron concentrations were estimated using PISCES2 model (Table S1a). In all cases, the abundance and expression of *ISIP* genes were normalized by the total diatom unigene abundance and expression, respectively, and were then scaled to the unit interval. The 3D plot shown below is derived from the three 2D scatter plots, with the color gradient representing the third dimension. The data were clustered using density clustering algorithms, resulting in a group of *Tara* Oceans sampling sites in which *ISIP* was only increased in metagenomics data (denoted Group 0 stations (40 stations; circles)) and others in which both metagenomic and metatranscriptomic data showed increases in *ISIP* levels (denoted Group 1 stations; 21 stations; triangles). The values corresponding to *Tara* Oceans stations in the Marquesas archipelago are labeled (122 - 125).

Tara Oceans sampling sites are colored according to the ocean region in the 3D plot: NPO, North Pacific Ocean; SO, Southern Ocean; SPO, South Pacific Ocean; NAO, North Atlantic Ocean; SAO, South Atlantic Ocean, MS, Mediterranean Sea; IO, Indian Ocean. (b) Relative abundance (above) and expression (below) of *ISIP* genes assigned at different levels of resolution in a diatom phylogenetic tree. The color code corresponds to the two clusters of stations defined in panel B based on *ISIP* patterns (red for Group 0 with variations only at metagenome levels; blue for Group 1 with variations in both metagenome and metatranscriptome levels).

Figure 6: Comparison of iron-driven changes in diatom *ISIP* gene abundance and expression and in the picocyanobacterial community from surface waters. Histograms of cell abundance of *Synechococcus* and *Prochlororococcus* clades at each *Tara* Oceans station are displayed, with stations sorted by hierarchical clustering of a Bray-Curtis distance matrix. The left panels indicate iron concentration estimates from PISCES2 model, and metagenome and metatranscriptome levels of diatom *ISIP* genes, including the resulting cluster type (circles and triangles as described in Figure 5).

Figure 7: The Marquesas study site, showing sampling sites, surface chlorophyll concentrations, and the local dynamics of the four eukaryotic IAAs. (a) Map of surface chlorophyll in the Marquesas area. Drifter and Provbio trajectories are indicated as well as the *Tara* Oceans sampling stations, with a zoom on Stations TARA_123 and TARA_124. For further details see main text and Supplementary Material S2. (b) Analysis of the dynamics of the four IAAs at the Marquesas archipelago stations in relation to their eigenlineage values (upper), richness (middle) and relative abundance (lower). All modules show negative eigenlineage values, with the exception of the DarkRed IAA. The DarkRed module positive eigenlineage scores significantly decrease within the bloom stations. The mean IAA relative abundance calculated over the global *Tara* Oceans dataset was subtracted from IAA relative abundance calculated at the Marquesas Islands. The increase in DarkRed relative abundance in station TARA_124 was due to a single *Prasinophycae* OTU. The mean IAA richness calculated over the global *Tara* Oceans dataset was subtracted from IAA richness calculated at the Marquesas Islands. Data indicates that the DarkRed IAA retains ca. 60% of its OTUs in low iron conditions,

a percentage that decreases in the bloom stations. (c) Relative abundance changes at the Marquesas Islands stations for IAA photosynthetic lineages with high iron correlation. The graph shows the list of IAA autotroph lineages with the highest statistically significant correlations against PISCES2 iron estimates ($p<0.05$) at a global scale, with the corresponding Pearson correlation coefficient, and their relative abundance at the Marquesas Islands sampling sites.

Figure 8: Variations in plankton community composition at Marquesas sampling sites. (a) Relative contribution of different autotrophic lineages to the total chlorophyll concentration in the euphotic zone (0 - 120 m), derived from photosynthetic pigment analysis and expressed as percent of the total measured chlorophyll. (b) Depth-integrated biomass ($mg \cdot C \cdot m^{-2}$) of autotrophs and mesozooplankton (> 300 µm) in the euphotic zone (0 – 120 m). (c) Relative abundance of *Prochlorococcus* and *Synechococcus* picocyanobacteria expressed as percent of the total *Prochlorococcus* plus *Synechococcus* abundance estimated from flow cytometry data. Genetic markers (*petB*) showed exactly the same trends (Supplementary Material S3). (d) Relative abundance (%) of ribotypes (18S-V9 tags) assigned to autotrophic eukaryote lineages at the surface (5 meters depth). Abundances were computed for the two size fractions containing the majority of autotrophic lineages, namely 0.8 - 5 µm and 5 – 20 µm size fractions. (e) Relative abundance (%) of ribotypes (18S-V9 Tags) assigned to metazoan lineages at the surface (5 m depth). Abundances were computed for the two size fractions containing the majority of metazoans, namely the 20 – 180 µm and 180 – 2000 µm size fractions. (f) Richness and diversity (exponential Shannon index) of eukaryotic autotrophs in two different size fractions estimated from the metabarcode data.

Figure 9: Variations in gene abundance and expression in cyanobacteria and diatoms at Marquesas sampling sites. (a-b) Differential expression patterns of iron-related genes from cyanobacteria *Prochlorococcus* (a) and *Synechococcus* (b) at stations TARA_122-125. Transcription values were normalized over genomic occurrence and are expressed relative to the levels observed at Station TARA_122 (index 100). The flavodoxin/ferredoxin ratio is also plotted (PetF/IsiB). (c) Relative abundances and mRNA levels of diatom genes potentially responsive to iron in metagenome and metatranscriptome datasets from Stations TARA_122 and 123. Values were normalized by total abundance or expression of all unigenes assigned to the

corresponding taxonomic group (*Pseudo-nitzschia* and *Thalassiosira*). For clarity we focused only on changes in 5-20 µm size fractions. Colors indicate the contribution of each station to the total levels. (d) Relative ratios between pairs of genes whose presence in the genome or transcriptional activity has been reported previously to be potentially responsive to iron bioavailability. For clarity, ferritin levels have been multiplied by a factor of 10 to be comparable with ISIP levels, and only 5-20 µm size fractions from Stations TARA_122 and 123 are compared.

Figure 1.

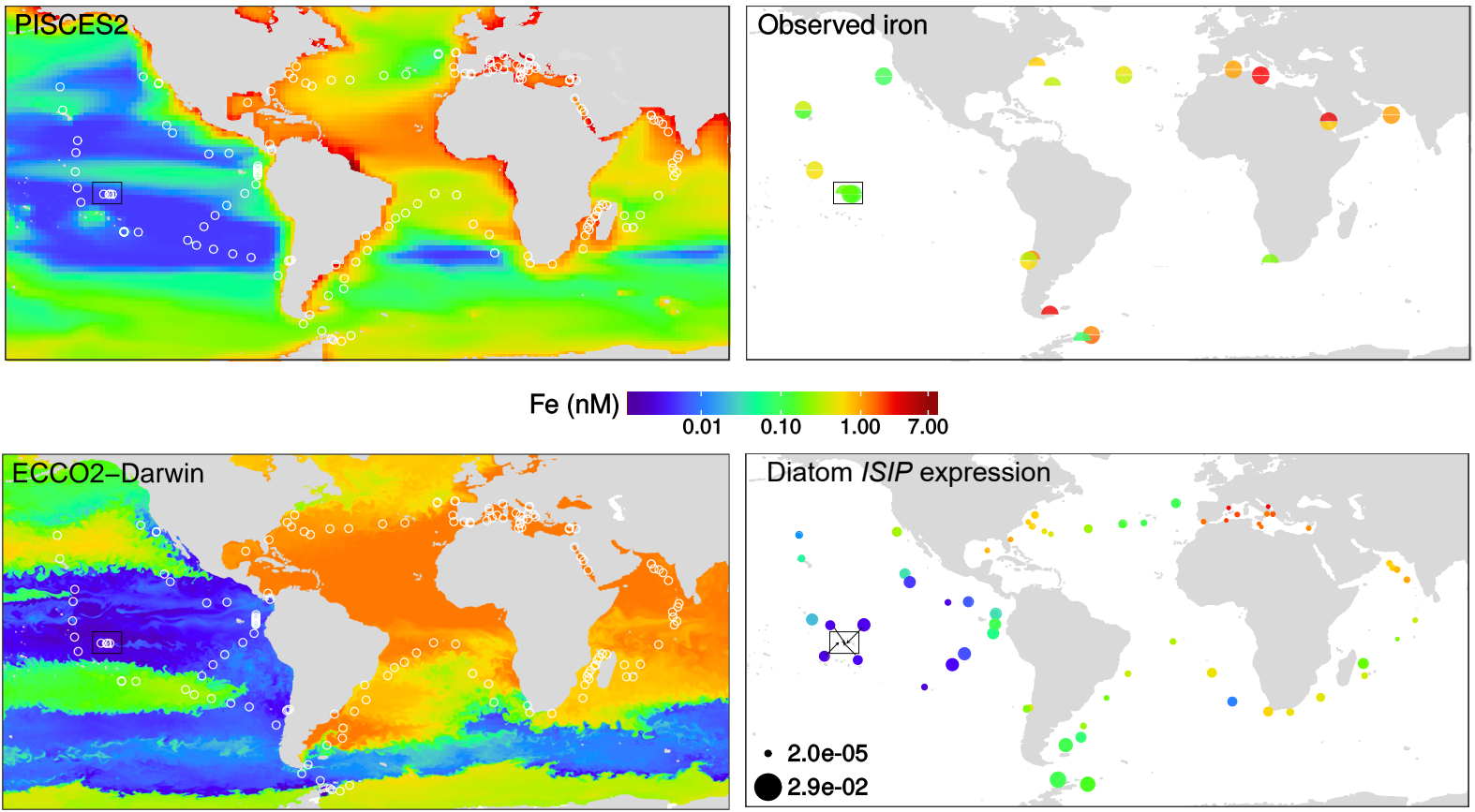

PISCES2
Observed iron
Fe (nM)
0.01
0.10
1.00
7.00
ECCO2–Darwin
Diatom *ISIP* expression
2.0e-05
2.9e-02

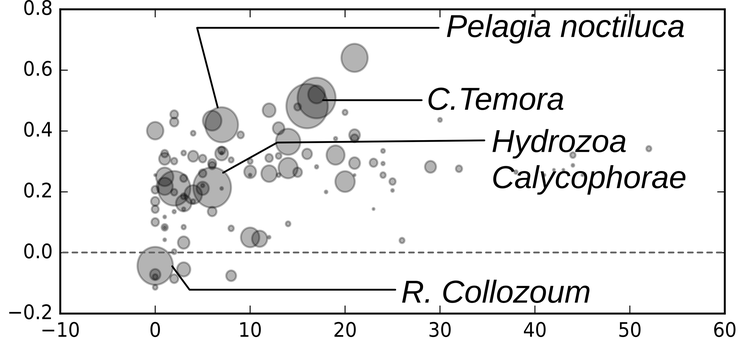
Figure 2.

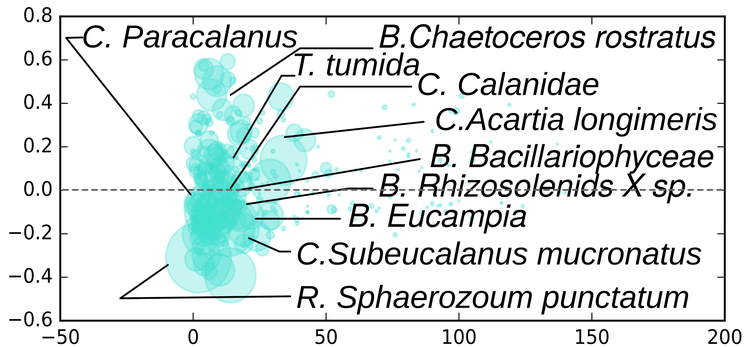
a

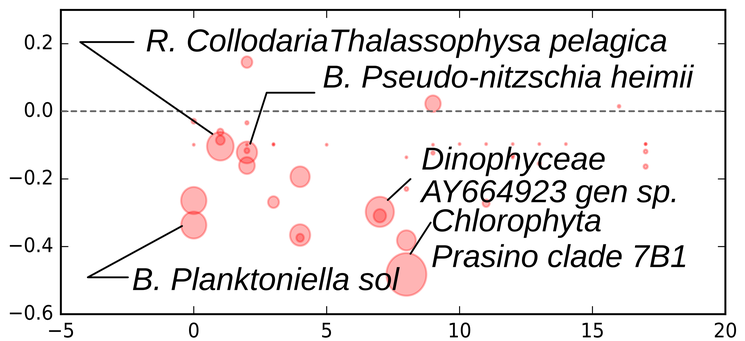
VIP+

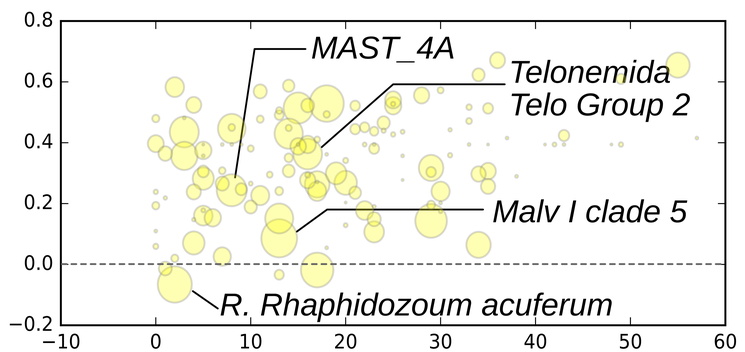
VIP-

Lineage correlation to iron



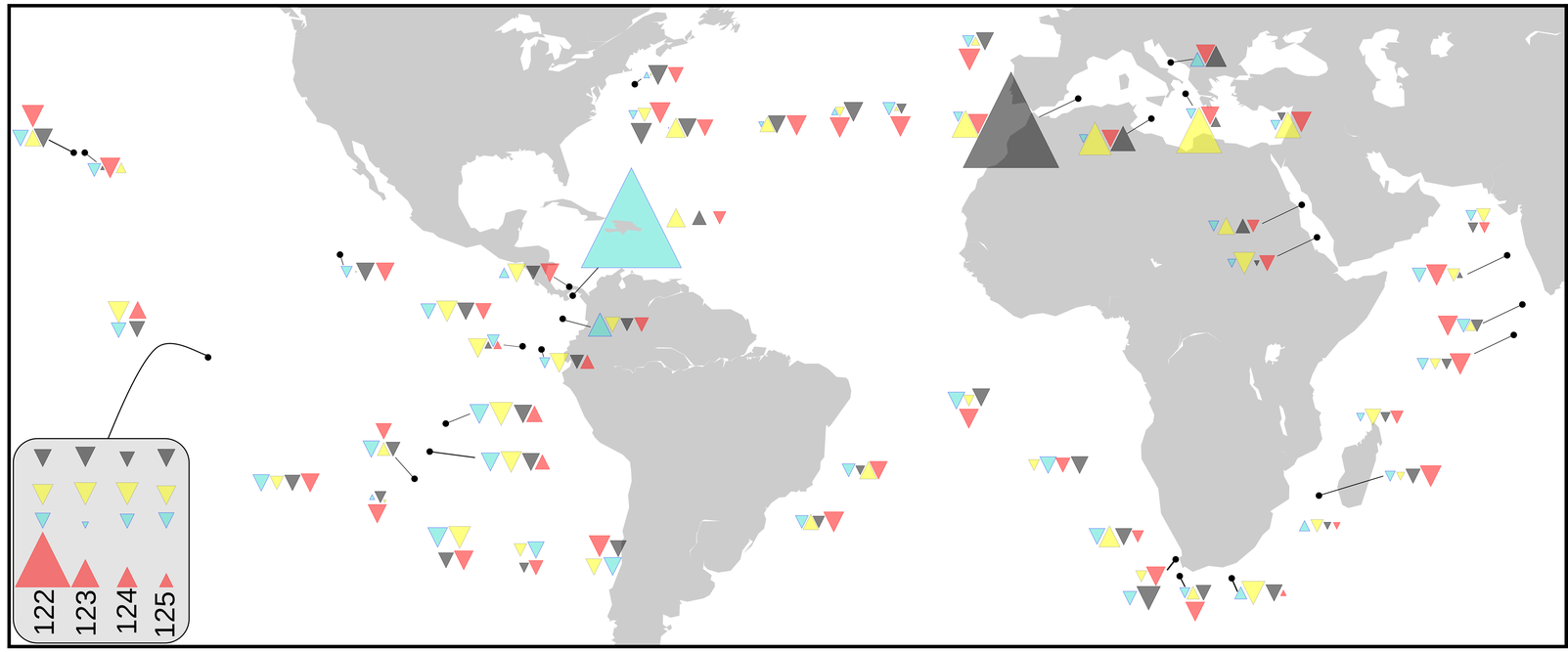


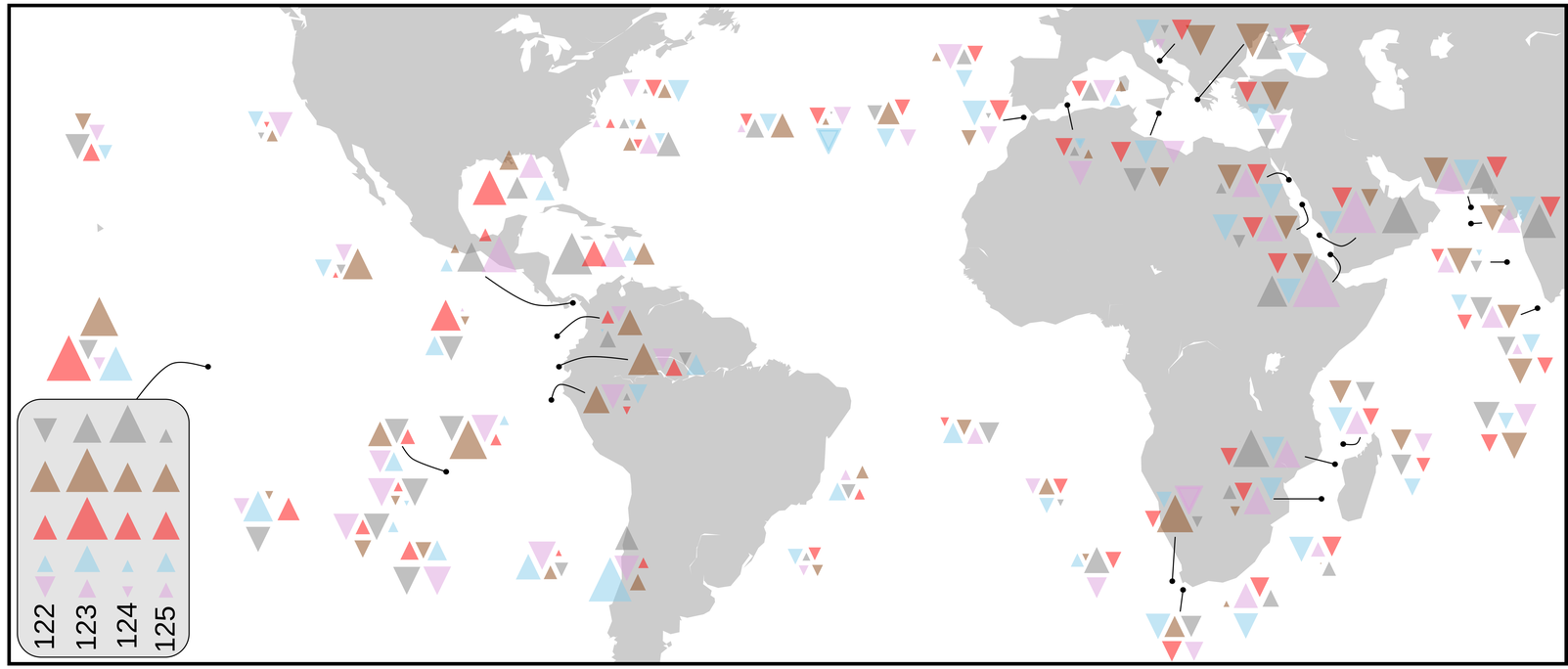




Lineage membership in subnetwork

b





Strong Positive / Negative eigenlineage values

Weak Positive / Negative eigenlineage values

Figure 3.

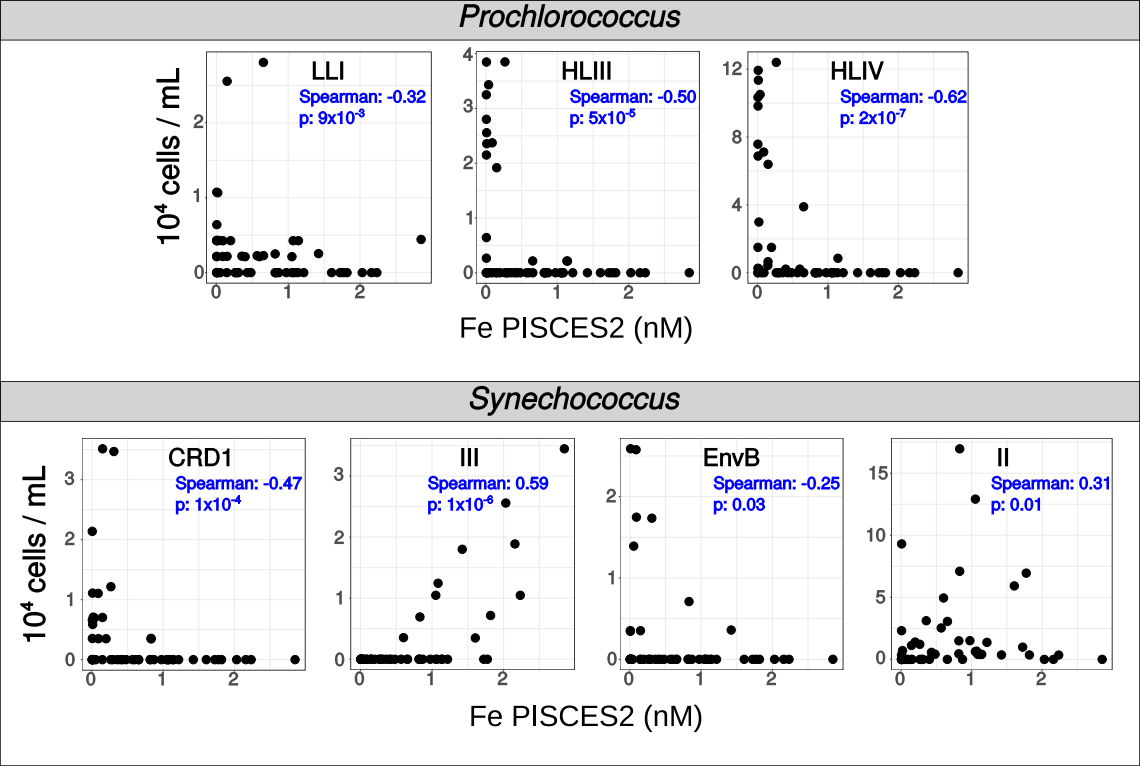

Prochlorococcus
LLI
Spearman: -0.32
p: 9x10-3
HLIII
Spearman: -0.50
p: 5x10-6
HLIV
Spearman: -0.62
p: 2x10-7
10^4 cells / mL
Fe PISCES2 (nM)
Synechococcus
CRD1
Spearman: -0.47
p: 1x10-4
III
Spearman: 0.59
p: 1x10-6
EnvB
Spearman: -0.25
p: 0.03
II
Spearman: 0.31
p: 0.01
10^4 cells / mL
Fe PISCES2 (nM)

Figure 4.

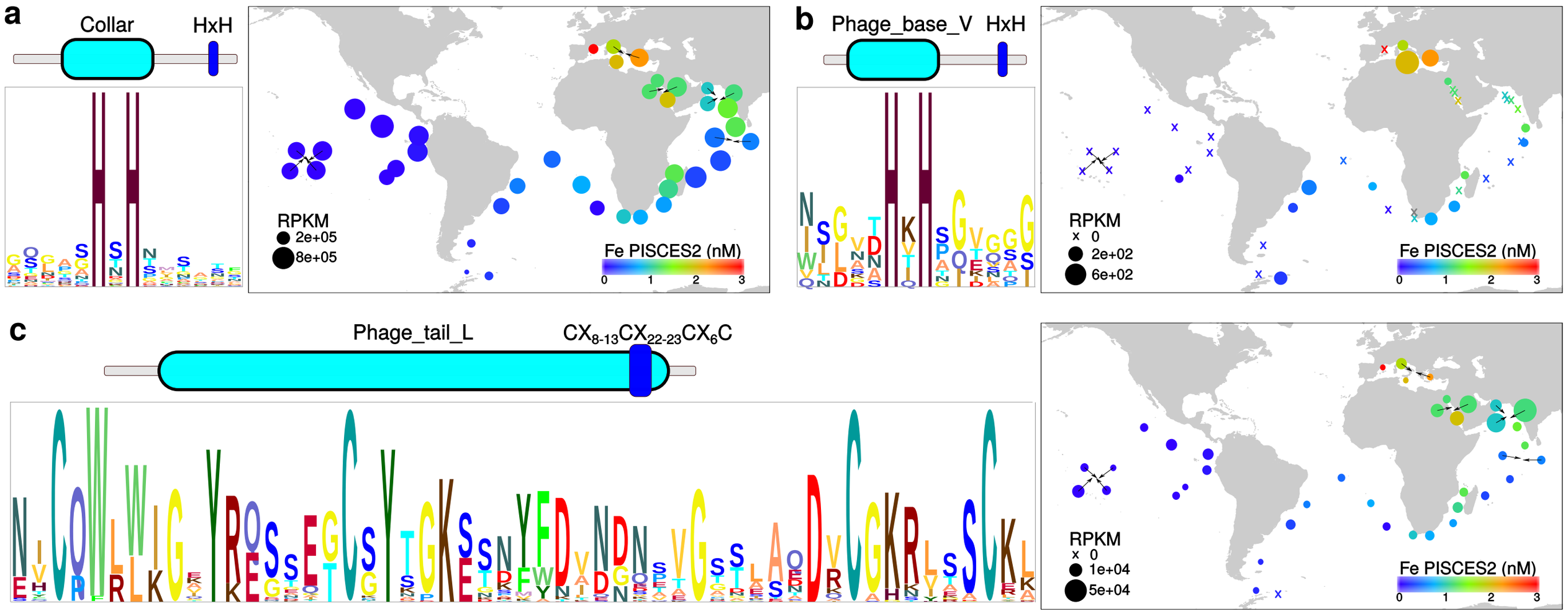

a
Collar
HxH
RPKM
2e+05
8e+05
Fe PISCES2 (nM)
0
1
2
3
b
Phage_base_V
HxH
RPKM
0
2e+02
6e+02
Fe PISCES2 (nM)
0
1
2
3
c
Phage_tail_L
CX8-13CX22-23CX6C
RPKM
0
1e+04
5e+04
Fe PISCES2 (nM)
0
1
2
3

Figure 5.

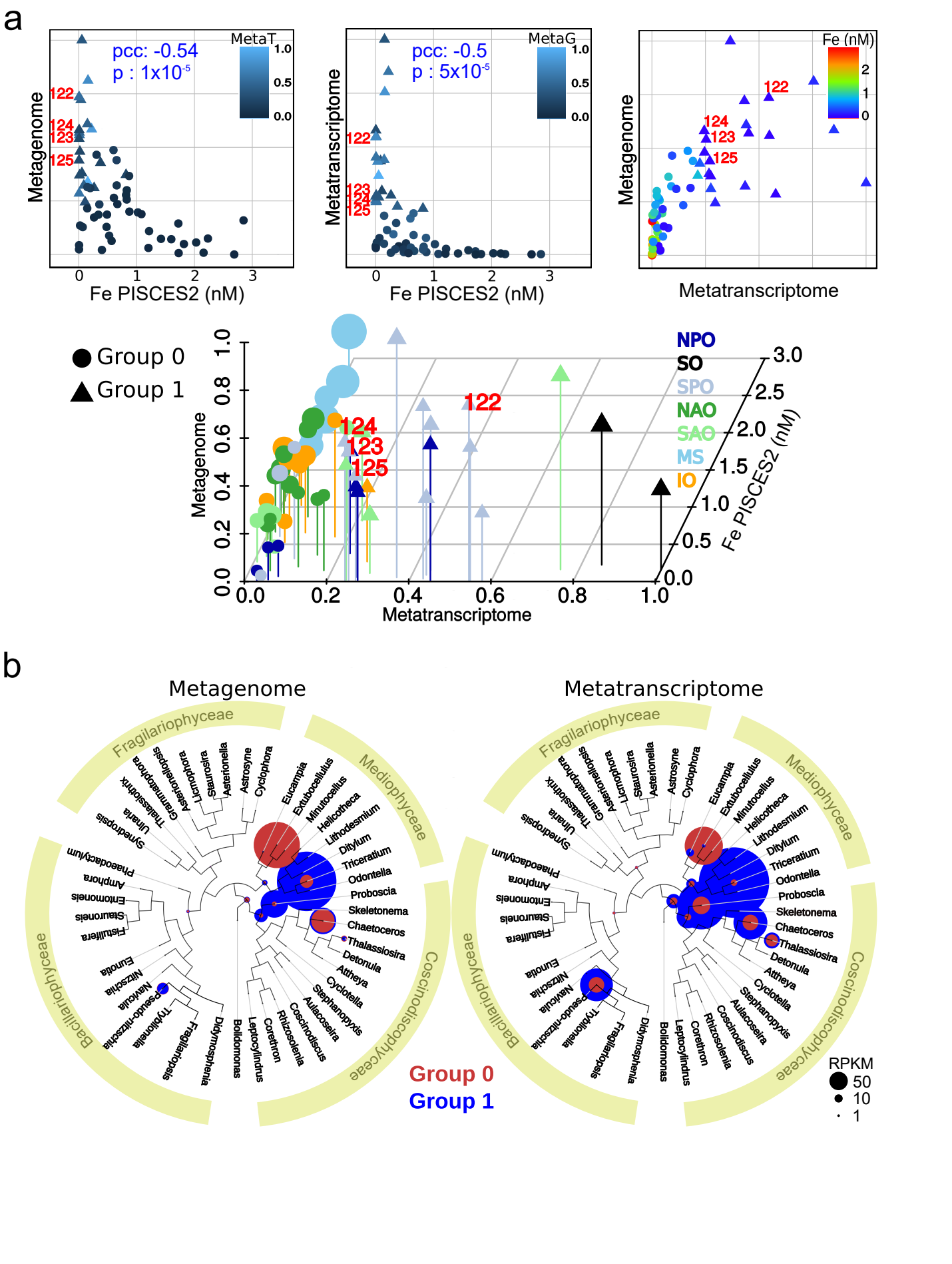

a
pcc: -0.54
p : 1x10-5
MetaT
Metagenome
Fe PISCES2 (nM)
pcc: -0.5
p : 5x10-5
MetaG
Metatranscriptome
Fe (nM)
122
123
124
125
Group 0
Group 1
NPO
SO
SPO
NAO
SAO
MS
IO
b
Metagenome
Metatranscriptome
Fragilariophyceae
Mediophyceae
Bacillariophyceae
Coscinodiscophyceae
Eucampia
Extubocellulus
Minutocellus
Helicotheca
Lithodesmium
Ditylum
Triceratium
Odontella
Proboscia
Skeletonema
Chaetoceros
Thalassiosira
Detonula
Attheya
Cyclotella
Stephanopyxis
Aulacoseira
Cosinodiscus
Rhizosolenia
Corethron
Leptocylindrus
Bolidomonas
Didymosphenia
Fragilariopsis
Pseudo-nitzschia
Nitzschia
Navicula
Entomoneis
Amphora
Phaeodactylum
Synedropsis
Ulnaria
Thalassiothrix
Grammatophora
Asterionellopsis
Lioloma
Asterionella
Astrosyne
Cyclophora
RPKM
50
10
1
Group 0
Group 1

Figure 6.

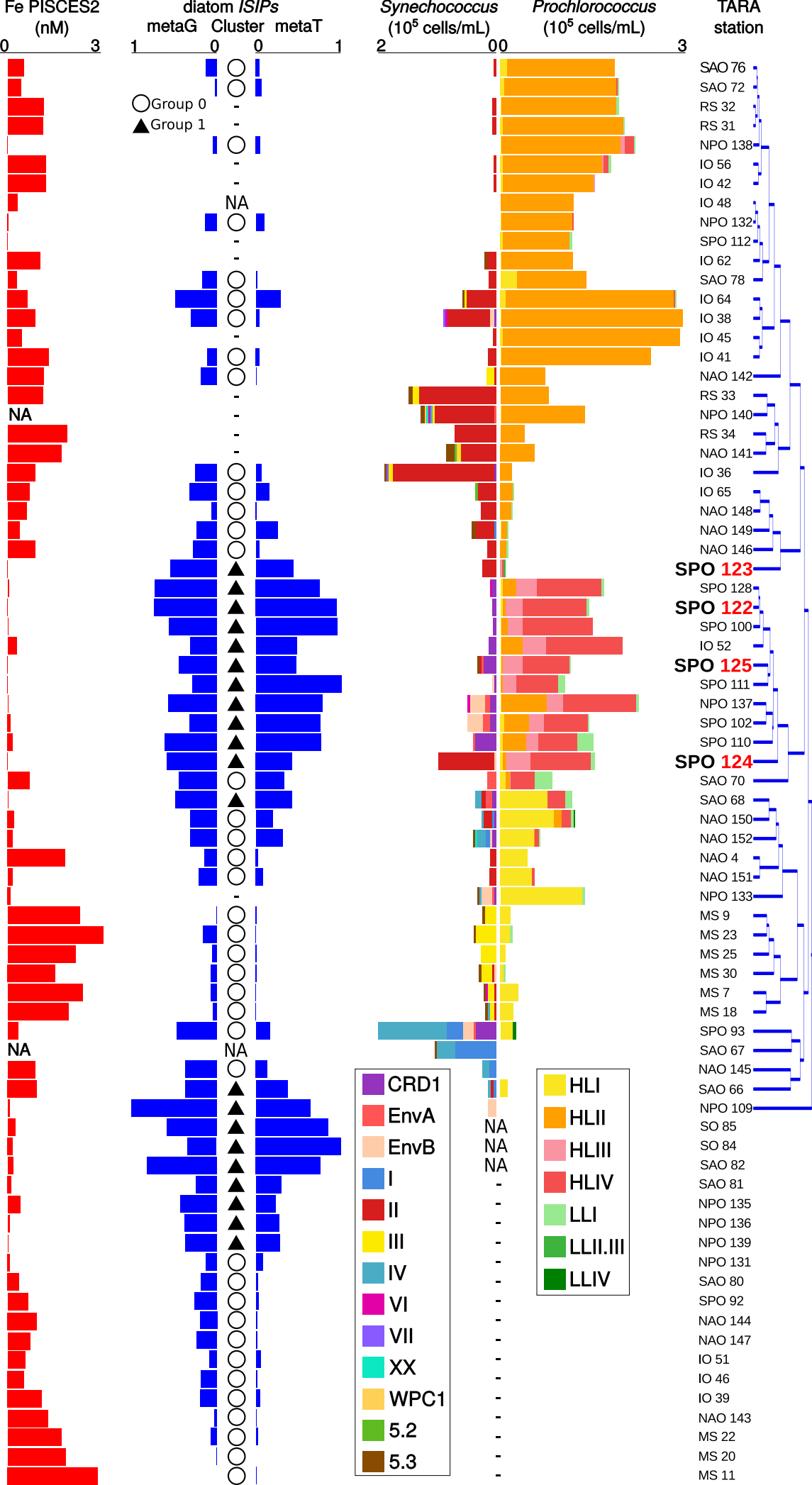

Fe PISCES2 (nM)
0
3
diatom ISIPs
metaG Cluster metaT
1 0 0 1
Group 0
Group 1
NA
Synechococcus (10^5 cells/mL)
2 0 0
Prochlorococcus (10^5 cells/mL)
3
TARA station
SAO 76
SAO 72
RS 32
RS 31
NPO 138
IO 56
IO 42
IO 48
NPO 132
SPO 112
IO 62
SAO 78
IO 64
IO 38
IO 45
IO 41
NAO 142
RS 33
NPO 140
RS 34
NAO 141
IO 36
IO 65
NAO 148
NAO 149
NAO 146
SPO 123
SPO 128
SPO 122
SPO 100
IO 52
SPO 125
SPO 111
NPO 137
SPO 102
SPO 110
SPO 124
SAO 70
SAO 68
NAO 150
NAO 152
NAO 4
NAO 151
NPO 133
MS 9
MS 23
MS 25
MS 30
MS 7
MS 18
SPO 93
SAO 67
NAO 145
SAO 66
NPO 109
SO 85
SO 84
SAO 82
SAO 81
NPO 135
NPO 136
NPO 139
NPO 131
SAO 80
SPO 92
NAO 144
NAO 147
IO 51
IO 46
IO 39
NAO 143
MS 22
MS 20
MS 11
CRD1
EnvA
EnvB
I
II
III
IV
VI
VII
XX
WPC1
5.2
5.3
HLI
HLII
HLIII
HLIV
LLI
LLII.III
LLIV

Figure 7.

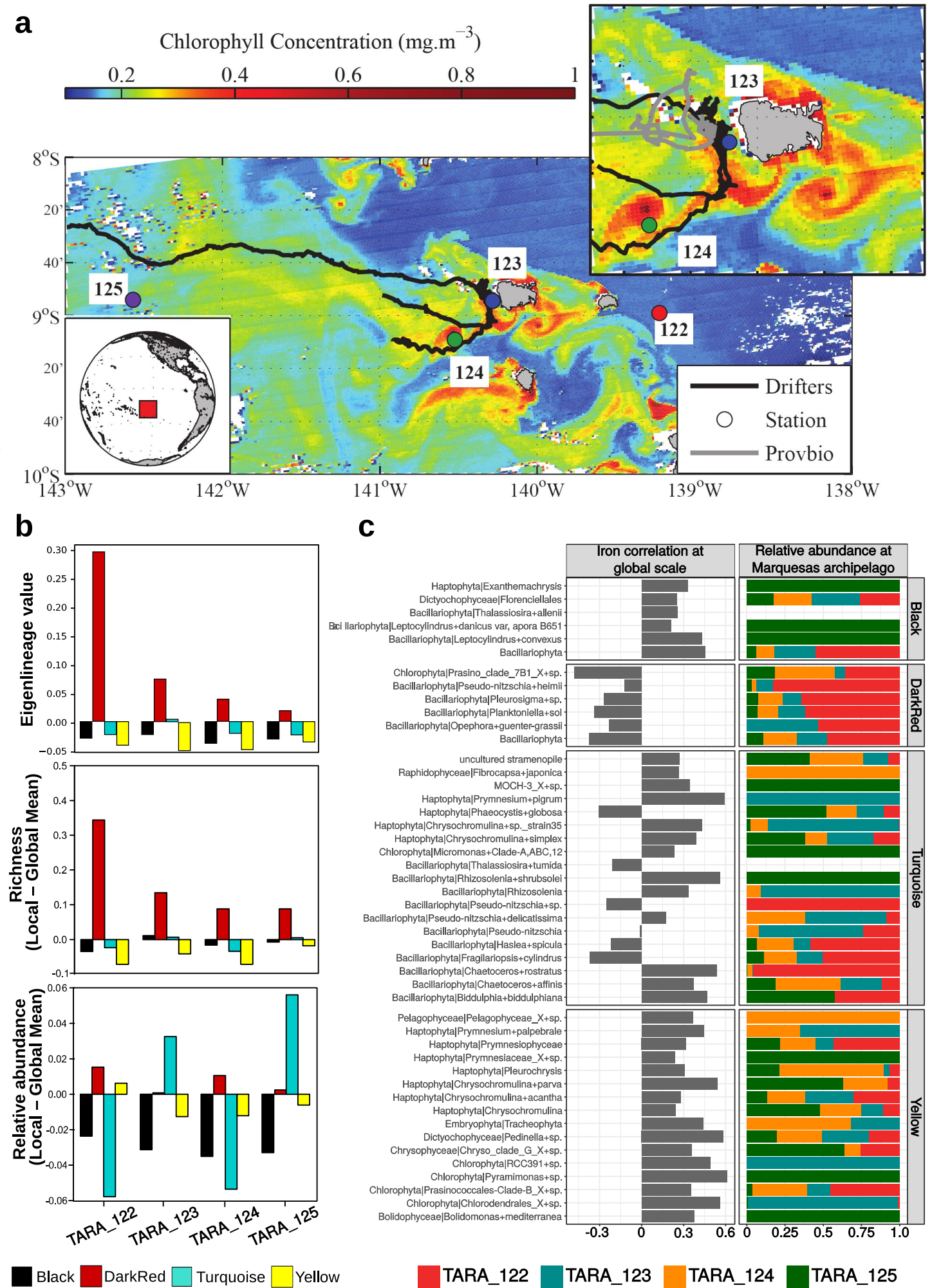

a
Chlorophyll Concentration (mg.m^{-3})
0.2
0.4
0.6
0.8
1
8°S
20'
40'
9°S
20'
40'
10°S
143°W
142°W
141°W
140°W
139°W
138°W
123
124
125
122
Drifters
Station
Provbio
b
Eigenlineage value
Richness (Local – Global Mean)
Relative abundance (Local – Global Mean)
TARA_122
TARA_123
TARA_124
TARA_125
Black
DarkRed
Turquoise
Yellow
c
Iron correlation at global scale
Relative abundance at Marquesas archipelago
Haptophyta|Exanthemachrysis
Dictyochophyceae|Florenciellales
Bacillariophyta|Thalassiosira+allenii
Bacillariophyta|Leptocylindrus+danicus var, apora B651
Bacillariophyta|Leptocylindrus+convexus
Bacillariophyta
Chlorophyta|Prasino_clade_7B1_X+sp.
Bacillariophyta|Pseudo-nitzschia+heimii
Bacillariophyta|Pleurosigma+sp.
Bacillariophyta|Planktoniella+sol
Bacillariophyta|Opephora+guenter-grassii
Bacillariophyta
uncultured stramenopile
Raphidophyceae|Fibrocapsa+japonica
MOCH-3_X+sp.
Haptophyta|Prymnesium+pigrum
Haptophyta|Phaeocystis+globosa
Haptophyta|Chrysochromulina+sp._strain35
Haptophyta|Chrysochromulina+simplex
Chlorophyta|Micromonas+Clade-A,ABC,12
Bacillariophyta|Thalassiosira+tumida
Bacillariophyta|Rhizosolenia+shrubsolei
Bacillariophyta|Rhizosolenia
Bacillariophyta|Pseudo-nitzschia+sp.
Bacillariophyta|Pseudo-nitzschia+delicatissima
Bacillariophyta|Pseudo-nitzschia
Bacillariophyta|Haslea+spicula
Bacillariophyta|Fragilariopsis+cylindrus
Bacillariophyta|Chaetoceros+rostratus
Bacillariophyta|Chaetoceros+affinis
Bacillariophyta|Biddulphia+biddulphiana
Pelagophyceae|Pelagophyceae_X+sp.
Haptophyta|Prymnesium+palpebrale
Haptophyta|Prymnesiophyceae
Haptophyta|Prymnesiaceae_X+sp.
Haptophyta|Pleurochrysis
Haptophyta|Chrysochromulina+parva
Haptophyta|Chrysochromulina+acantha
Haptophyta|Chrysochromulina
Embryophyta|Tracheophyta
Dictyochophyceae|Pedinella+sp.
Chrysophyceae|Chryso_clade_G_X+sp.
Chlorophyta|RCC391+sp.
Chlorophyta|Pyramimonas+sp.
Chlorophyta|Prasinococcales-Clade-B_X+sp.
Chlorophyta|Chlorodendrales_X+sp.
Bolidophyceae|Bolidomonas+mediterranea
Black
DarkRed
Turquoise
Yellow
TARA_122
TARA_123
TARA_124
TARA_125

Figure 8.

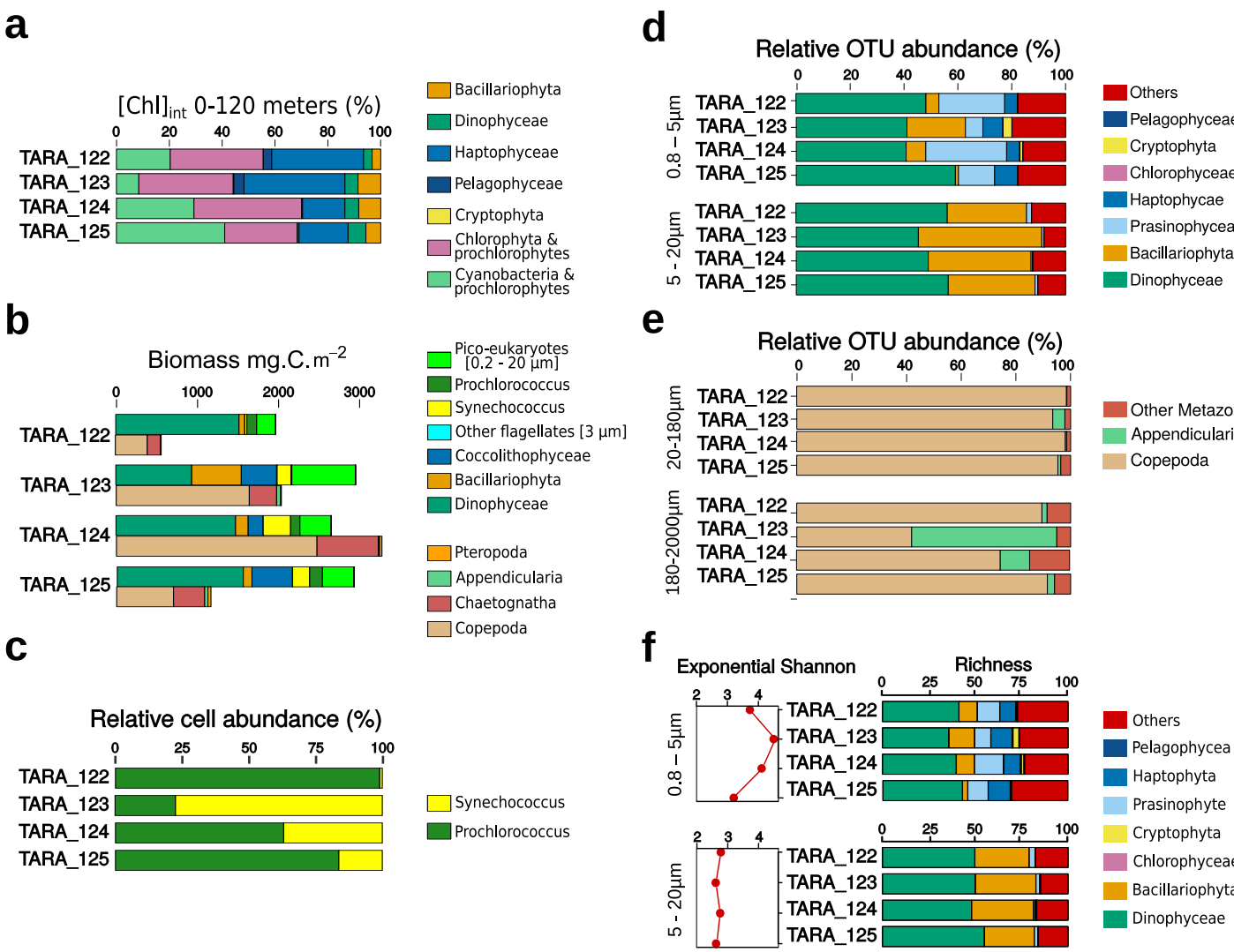
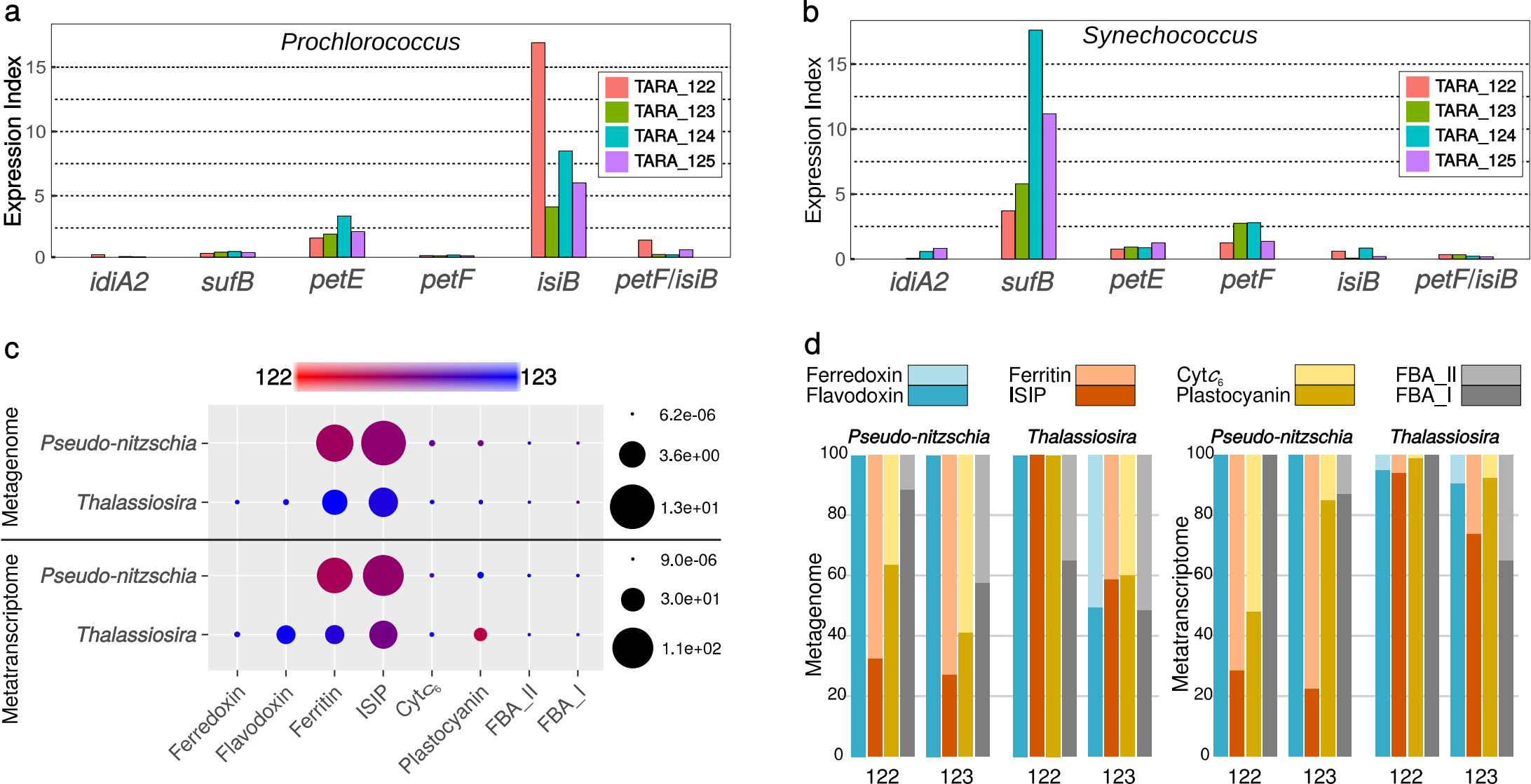

a
[Chl]int 0-120 meters (%)
0
20
40
60
80
100
TARA_122
TARA_123
TARA_124
TARA_125
Bacillariophyta
Dinophyceae
Haptophyceae
Pelagophyceae
Cryptophyta
Chlorophyta & prochlorophytes
Cyanobacteria & prochlorophytes
b
Biomass mg.C.m−2
0
1000
2000
3000
TARA_122
TARA_123
TARA_124
TARA_125
Pico-eukaryotes [0.2 - 20 µm]
Prochlorococcus
Synechococcus
Other flagellates [3 µm]
Coccolithophyceae
Bacillariophyta
Dinophyceae
Pteropoda
Appendicularia
Chaetognatha
Copepoda
c
Relative cell abundance (%)
0
25
50
75
100
TARA_122
TARA_123
TARA_124
TARA_125
Synechococcus
Prochlorococcus
d
Relative OTU abundance (%)
0
20
40
60
80
100
0.8 – 5µm
5 - 20µm
TARA_122
TARA_123
TARA_124
TARA_125
Others
Pelagophyceae
Cryptophyta
Chlorophyceae
Haptophycae
Prasinophyceae
Bacillariophyta
Dinophyceae
e
Relative OTU abundance (%)
0
20
40
60
80
100
20-180µm
180-2000µm
TARA_122
TARA_123
TARA_124
TARA_125
Other Metazoa
Appendicularia
Copepoda
f
Exponential Shannon
2
3
4
Richness
0
25
50
75
100
0.8 – 5µm
5 - 20µm
TARA_122
TARA_123
TARA_124
TARA_125
Others
Pelagophycea
Haptophyta
Prasinophyte
Cryptophyta
Chlorophyceae
Bacillariophyta
Dinophyceae

Figure 9.


a
Prochlorococcus
Expression Index
0
5
10
15
TARA_122
TARA_123
TARA_124
TARA_125
idiA2
sufB
petE
petF
isiB
petF/isiB
b
Synechococcus
Expression Index
0
5
10
15
TARA_122
TARA_123
TARA_124
TARA_125
idiA2
sufB
petE
petF
isiB
petF/isiB
c
122
123
Metagenome
Metatranscriptome
Pseudo-nitzschia
Thalassiosira
Pseudo-nitzschia
Thalassiosira
6.2e-06
3.6e+00
1.3e+01
9.0e-06
3.0e+01
1.1e+02
Ferredoxin
Flavodoxin
Ferritin
ISIP
Cytc6
Plastocyanin
FBA_II
FBA_I
d
Ferredoxin
Flavodoxin
Ferritin
ISIP
Cytc6
Plastocyanin
FBA_II
FBA_I
Pseudo-nitzschia
Thalassiosira
Pseudo-nitzschia
Thalassiosira
Metagenome
Metatranscriptome
0
20
40
60
80
100
122
123
122
123
122
123
122
123